



DAGSTUHL REPORTS

Volume 3, Issue 11, November 2013

Computational Audio Analysis (Dagstuhl Seminar 13451) <i>Meinard Müller, Shrikanth S. Narayanan, and Björn Schuller</i>	1
Proxemics in Human-Computer Interaction (Dagstuhl Seminar 13452) <i>Saul Greenberg, Kasper Hornbæk, Aaron Quigley, Harald Reiterer, and Roman Rädle</i>	29
Electronic Markets and Auctions (Dagstuhl Seminar 13461) <i>Yishay Mansour, Benny Moldovanu, Noam Nisan, and Berthold Vöcking</i>	58
Computational Models of Language Meaning in Context (Dagstuhl Seminar 13462) <i>Hans Kamp, Alessandro Lenci, and James Pustejovsky</i>	79
Synchronous Programming (Dagstuhl Seminar 13471) <i>Stephen A. Edwards, Alain Girault, and Klaus Schneider</i>	117
Global Measurement Framework (Dagstuhl Seminar 13472) <i>Philip Eardley, Marco Mellia, Jörg Ott, Jürgen Schönwälder, and Henning Schulzrinne</i>	144
Unleashing Operational Process Mining (Dagstuhl Seminar 13481) <i>Rafael Accorsi, Ernesto Damiani, and Wil van der Aalst</i>	154
Forensic Computing (Dagstuhl Seminar 13482) <i>Felix C. Freiling, Gerrit Hornung, and Radim Polcák</i>	193

ISSN 2192-5283

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany.

Online available at <http://www.dagstuhl.de/dagrep>

Publication date

March, 2014

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

License

This work is licensed under a Creative Commons Attribution 3.0 Unported license: CC-BY.



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Aims and Scope

The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.

In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
 - an overview of the talks given during the seminar (summarized as talk abstracts), and
 - summaries from working groups (if applicable).
- This basic framework can be extended by suitable contributions that are related to the program of the seminar, e.g. summaries from panel discussions or open problem sessions.

Editorial Board

- Susanne Albers
- Bernd Becker
- Karsten Berns
- Stephan Diehl
- Hannes Hartenstein
- Stephan Merz
- Bernhard Mitschang
- Bernhard Nebel
- Han La Poutré
- Bernt Schiele
- Nicole Schweikardt
- Raimund Seidel
- Michael Waidner
- Reinhard Wilhelm (*Editor-in-Chief*)

Editorial Office

Marc Herbstritt (*Managing Editor*)

Jutka Gasirowski (*Editorial Assistance*)

Thomas Schillo (*Technical Assistance*)

Contact

Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de

Digital Object Identifier: 10.4230/DagRep.3.11.i

www.dagstuhl.de/dagrep

Computational Audio Analysis

Edited by

Meinard Müller¹, Shrikanth S. Narayanan², and Björn Schuller³

- 1 International Audio Laboratories Erlangen,
Universität Erlangen-Nürnberg, DE, meinard.mueller@audiolabs-erlangen.de
- 2 USC – Los Angeles, US, shri@sipi.usc.edu
- 3 TU München, DE, schuller@tum.de

Abstract

Compared to traditional speech, music, or sound processing, the computational analysis of general audio data has a relatively young research history. In particular, the extraction of affective information (i. e., information that does not deal with the ‘immediate’ nature of the content such as the spoken words or note events) from audio signals has become an important research strand with a huge increase of interest in academia and industry. At an early stage of this novel research direction, many analysis techniques and representations were simply transferred from the speech domain to other audio domains. However, general audio signals (including their affective aspects) typically possess acoustic and structural characteristics that distinguish them from spoken language or isolated ‘controlled’ music or sound events. In the Dagstuhl Seminar 13451 titled “Computational Audio Analysis” we discussed the development of novel machine learning as well as signal processing techniques that are applicable for a wide range of audio signals and analysis tasks. In particular, we looked at a variety of sounds besides speech such as music recordings, animal sounds, environmental sounds, and mixtures thereof. In this report, we give an overview of the various contributions and results of the seminar. We start with an executive summary, which describes the main topics, goals, and group activities. Then, one finds a list of abstracts giving a more detailed overview of the participants’ contributions as well as of the ideas and results discussed in the group meetings of our seminar. To conclude, an attempt is made to define the field as given by the views of the participants.

Seminar 03.–08. November, 2013 – www.dagstuhl.de/13451

1998 ACM Subject Classification I.5.4 Applications—Waveform analysis, H.5.5 Sound and Music Computing, H.1.2 User/Machine Systems, H.5.1 Multimedia Information Systems

Keywords and phrases Audio Analysis, Signal Processing, Machine Learning, Sound, Speech, Music, Affective Computing

Digital Object Identifier 10.4230/DagRep.3.11.1

Edited in cooperation with Jonathan Driedger

1 Executive Summary

Meinard Müller

Shrikanth S. Narayanan

Björn Schuller

License © Creative Commons BY 3.0 Unported license
© Meinard Müller, Shrikanth S. Narayanan, and Björn Schuller

With the rapid growth and omnipresence of digitized multimedia data, the processing, analysis, and understanding of such data by means of automated methods has become a



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Computational Audio Analysis, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 1–28

Editors: Meinard Müller, Shrikanth S. Narayanan, and Björn Schuller



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

central issue in computer science and associated areas of research. As for the acoustic domain, audio analysis has traditionally been focused on data related to speech with the goal to recognize and transcribe the spoken words. In this seminar, we considered current and future audio analysis tasks that go beyond the classical speech recognition scenario. For example, we looked at the computational analysis of speech with regard to the speakers' traits (e. g., gender, age, height, cultural and social background), physical conditions (e. g., sleepiness, alcohol intoxication, health state), or emotion-related and affective states (e. g., stress, interest, confidence, frustration). So, rather than recognizing *what* is being said, the goal is to find out *how* and *by whom* it is being said. Besides speech, there is a rich variety of sounds such as music recordings, animal sounds, environmental sounds, and combinations thereof. Just as for the speech domain, we discussed how to decompose and classify the content of complex sound mixtures with the objective to infer semantically meaningful information.

When dealing with specific audio domains such as speech or music, it is crucial to properly understand and apply the appropriate domain-specific properties, be they acoustic, linguistic, or musical. Furthermore, data-driven learning techniques that exploit the availability of carefully annotated audio material have successfully been used for recognition and classification tasks. In this seminar, we discussed issues that arise when dealing with rather vague categories as in emotion recognition or when considering general audio sources such as environmental sounds. In such scenarios, model assumptions are often violated, or it becomes impossible to define explicit representations or models. Furthermore, for non-standard audio material, annotated datasets are hardly available. Also, data-driven methods that are used in speech recognition are (often) not directly applicable in this context; instead semi-supervised or unsupervised learning techniques can be a promising approach to remedy these issues. Another central topic of this seminar was concerned with the problem of source separation. In the real world, acoustic data is very complex typically consisting of a superposition of overlapping speech, music, and general sound sources. Therefore, efficient source separation techniques are required that allow for splitting up, re-synthesizing, analyzing, and classifying the individual sources—a problem that, for general audio signals, is yet not well understood.

In this executive summary, we give a short overview of the main topics addressed in this seminar. We start by briefly describing the background of the participants and the overall organization. We then give an overview of the presentations of the participants and the results obtained from the different working groups. Finally, we reflect on the most important aspects of this seminar and conclude with future implications.

Participants, Interaction, Activities

In our seminar, we had 41 participants, who came from various countries around the world including North America (10 participants), Japan (1 participant), and Europe (Austria, Belgium, Finland, France, Germany, Greece, Italy, Netherlands, Spain, United Kingdom). Most of the participants came to Dagstuhl for the first time and expressed enthusiasm about the open and retreat-like atmosphere. Besides its international character, the seminar was also highly interdisciplinary. While most of the participating researchers are working in the fields of signal processing and machine learning, we have had participants with a background in cognition, human computer interaction, music, linguistics, and other fields. This made the seminar very special in having many cross-disciplinary intersections and provoking discussions as well as numerous social activities including common music making.

Overall Organization and Schedule

Dagstuhl seminars are known for having a high degree of flexibility and interactivity, which allow participants to discuss ideas and to raise questions rather than to present research results. Following this tradition, we fixed the schedule during the seminar asking for spontaneous contributions with future-oriented content, thus avoiding a conference-like atmosphere, where the focus is on past research achievements. The first two days were used to let people introduce themselves, present scientific problems they are particularly interested in and express their expectations and wishes for the seminar. In addition, we have had six initial stimulus talks, where specific participants were asked to address some burning questions on speech, music, and sound processing from a more meta point of view, see also Section 3. Rather than being usual presentations, most of these stimulus talks seamlessly moved towards an open discussion of the plenum. Based on this input, the second day concluded with a brainstorming session, where we identified central topics covering the participants' interests and discussed the schedule and format of the subsequent days. To discuss these topics, we split up into five groups, each group discussing one of the topics in greater depth in parallel sessions on Wednesday morning. The results and conclusions of these group meetings were then presented to the plenum on Thursday morning, which resulted in vivid discussions. Continuing the previous activities, further parallel group meetings were held on Thursday afternoon, the results of which being presented on Friday morning. Finally, asking each participant to give a short (written) statement of what he or she understands by the seminar's overall topic "Computational Audio Analysis," we had a very entertaining and stimulating session by going through and discussing all these statements one by one. The result of this session can be found in Section 6. In summary, having a mixture of different presentation styles and group meetings gave all participants the opportunity for presenting and discussing their ideas, while avoiding a monotonous conference-like atmosphere.

Main Topics

We discussed various topics that addressed the challenges when dealing with mixtures of general and non-standard acoustic data. A particular focus was put on data representations and analysis techniques including audio signal processing, machine learning, and probabilistic models. After a joint brainstorming session, we agreed on discussing five central topics which fitted in the overall theme of the seminar and reflected the participants' interests. We now give a brief summary of these topics, which were addressed in the parallel group meetings and resulting panel discussions. A more detailed summary of the outcome of the group sessions can be found in Section 4.

1. The "*Small Data*" group looked at audio analysis and classification scenarios where only few labeled examples or small amounts of (training) data are available. In such scenarios, machine learning techniques that depend on large amounts of (training) data ("Big Data") are not applicable. Various strategies including model-based as well as semi- and unsupervised approaches were discussed.
2. The "*Source Separation*" group addressed the task of decomposing a given sound mixture into elementary sources, which is not only a fundamental problem in audio processing, but also constitutes an intellectual and interdisciplinary challenge. Besides questioning the way the source separation problem is often posed, the need of concrete application scenarios as well as the objective of suitable evaluation metrics were discussed.

3. The “*Interaction and Affect*” group discussed the question on how to generate and interpret signals that express interactions between different agents. One main conclusion was that one requires more flexible models that better adapts to the temporal and situational context as well as to the agents’ roles, behaviors and traits.
4. The “*Knowledge Representation*” group addressed the issue of how knowledge can be used to define and derive sound units that can be used as elementary building blocks for a wide range of applications. Based on deep neural network techniques, the group discussed how database information and other meta-data can be better exploited and integrated using feed-forward as well as recurrent architectures.
5. The “*Unsupervised Learning*” group looked at the problem on how to learn the structure of data without reference to external objectives. Besides issues on learning meaningful elementary units, the need of considering hierarchies of abstractions and multi-layer characterizations was discussed.

Besides an extensive discussion of these five main topics, we have had many further contributions and smaller discussions on issues that concern natural human machine communication, human centered audio processing, computational paralinguistics, sound processing in everyday environments, acoustic monitoring, informed source separation, and audio structure analysis.

Conclusions

In our seminar, we addressed central issues on how to process audio material of various types and degrees of complexity. In view of the richness and multitude of acoustic data, one requires representations and machine learning techniques that allow for capturing and coupling various sources of information. Therefore, unsupervised and semi-supervised learning procedures are needed in scenarios where only very few examples and poor training resources are available. Also, source separation techniques are needed, which yield meaningful audio decomposition results even when having only limited knowledge on the type of audio. Another central issue of this seminar was how to bring in the human into the audio processing pipeline. On the one hand, we discussed how we can learn from the way human process and perceive sounds. On the other hand, we addressed the issue on extracting human-related parameters such as affective and paralinguistic information from sound sources. These discussions showed that understanding and processing complex sound mixtures using computational tools poses many challenging research problems yet to be solved.

The Dagstuhl seminar gave us the opportunity for discussing such issues in an inspiring and retreat-like atmosphere. The generation of novel, technically oriented scientific contributions was not the focus of the seminar. Naturally, many of the contributions and discussions were on a rather abstract level, laying the foundations for future projects and collaborations. Thus, the main impact of the seminar is likely to take place in the medium to long term. Some more immediate results, such as plans to share research data and software, also arose from the discussions. As measurable outputs from the seminar, we expect to see several joint papers and applications for funding. Beside the scientific aspect, the social aspect of our seminar was just as important. We had an interdisciplinary, international, and very interactive group of researchers, consisting of leaders and future leaders in our field. Most of our participants visited Dagstuhl for the first time and enthusiastically praised the open and inspiring atmosphere. The group dynamics were excellent with many personal exchanges and common activities. Some scientists mentioned their appreciation of having the opportunity

for prolonged discussions with researchers from neighboring research fields—something which is often impossible during conference-like events.

In conclusion, our expectations of the seminar were not only met but exceeded, in particular with respect to networking and community building. Last but not least, we heartily thank the Dagstuhl board for allowing us to organize this seminar, the Dagstuhl office for their great support in the organization process, and the entire Dagstuhl staff for their excellent services during the seminar.

2 Table of Contents

Executive Summary

<i>Meinard Müller, Shrikanth S. Narayanan, and Björn Schuller</i>	1
-----------------------------------------------------------------------------	---

Stimulus Talks

Multimedia Analysis for the Poor <i>Xavier Anguera</i>	8
Interpreting ‘Intentional’ Behaviour in Audio Scenes <i>Roger K. Moore</i>	9
Semantic-Affective Models for Multimodal Signal Processing <i>Alexandros Potamianos</i>	9
Exploitation of Human Perception Principles in Audio Processing Systems <i>Gaël Richard</i>	10
Stop Listening to Speech, Language, and Vision Research! <i>Paris Smaragdis</i>	10
Sound Event Detection and Recognition in Everyday Environments <i>Tuomas Virtanen</i>	11

Group Sessions

Small Data (Learning from Few Examples) <i>Bryan Pardo, Xavier Anguera, Jonathan Driedger, Bernd Edler, Jort Gemmeke, Franz Graf, Gernot Kubin, Frank Kurth, Meinard Müller, and Christian Uhle</i>	11
Interaction and Affect <i>Martin Heckmann, Murtaza Bulut, Carlos Busso, Nick Campbell, Laurence Devillers, Anna Esposito, Sungbok Lee, Roger Moore, Mark Sandler, Khiet Truong, and Rita Singh</i>	12
Learning of Units and Knowledge Representation <i>Florian Metze, Xavier Anguera, Sebastian Ewert, Jort Gemmeke, Dorothea Kolossa, Emily Mower Provost, Björn Schuller, and Joan Serra</i>	13
Unsupervised Learning for Audio <i>Tuomas Virtanen, Jon Barker, Shrikanth Narayanan, Alexandros Potamianos, Bhiksha Raj, Gaël Richard, Rita Singh, Paris Smaragdis, Stefano Squartini, and Shiva Sundaram</i>	14
Source Separation <i>Christian Uhle, Jonathan Driedger, Bernd Edler, Sebastian Ewert, Franz Graf, Gernot Kubin, Meinard Müller, Nobutaka Ono, Bryan Pardo, and Joan Serra</i>	15

Further Topics

Engineering Selective Attention into Acoustic Scene Analysis Systems. <i>Jon Barker</i>	15
Compensate Lexical/Speaker/Environment Variability for Speech Emotion Recognition <i>Carlos Busso</i>	16

Interpretation and Computational Audio Analysis <i>Laurence Devillers</i>	17
Perceptually Appealing Reconstruction of Spectrally Modified Signals <i>Jonathan Driedger and Meinard Müller</i>	17
The Situated Multimodal Facets of Human Communication <i>Anna Esposito</i>	17
Bayes and Beyond Bayes: The Integration of Prior Knowledge <i>Sebastian Ewert</i>	18
NMF meet Dynamics <i>Cédric Févotte</i>	19
Features beyond Machine Learning <i>Martin Heckmann</i>	20
Detection of Repeated Signal Components and Applications to Audio Analysis <i>Frank Kurth</i>	21
Informed Source Separation for Music Signals <i>Meinard Müller and Jonathan Driedger</i>	21
Approaching Cross-Audio Computer Audition <i>Björn Schuller</i>	22
What can we Learn from Massive Music Archives? <i>Joan Serra</i>	22
Acoustic Monitoring in Smart Home Environments: A Holistic Perspective <i>Stefano Squartini</i>	23
Sound Processing in Everyday Environments <i>Emmanuel Vincent</i>	24
Towards a Definition of Computational Audio Analysis (CAA)	24
Participants	28

3 Stimulus Talks

3.1 Multimedia Analysis for the Poor

Xavier Anguera (Telefónica Research – Barcelona, ES)

License  Creative Commons BY 3.0 Unported license
© Xavier Anguera

In many areas of multimedia analysis (i. e., when extracting knowledge from multimedia or multimodal data) one usually first derives models from corpora of annotated training data, which are then applied to some unknown data. Highly performing systems have been built using this methodology in the past. However, it must not be overlooked that producing high-quality annotated data for training takes time and resources, which are not always available. Examples of high-quality labeling scarcity can be seen when trying to analyze highly diverse data like what is found on online media sources such as YouTube or SoundCloud, or with rare languages in speech (i. e. those languages for which the number of speakers is too small to attract commercial interest). For this reason it becomes very relevant to explore new avenues to be able to extract knowledge with no (or very limited) labeled examples. There are already many efforts in this direction within the research community such as:

- Speech: Audio summarization through the analysis of repetitions in the audio stream; query-by-example spoken term detection; training systems (e. g., large vocabulary speech recognition) on little transcribed data, or on low quality transcripts (e.g. close captions) data.
- Music: Structural analysis of songs.
- Image processing: Unsupervised concept extraction (e. g., the system developed by Google and G. Hinton to automatically learn how to recognize cats in Youtube videos).
- Text: Unsupervised document clustering and topic detection.
- Bioinformatics: Unsupervised repetitions/structure detection and finding mutations.

In this inspirational talk, I first motivated the need to do research on unsupervised and semi-supervised algorithms to tackle problems like those mentioned above. Then, after presenting some examples of technologies that are able to perform well with these constraints, I described the task of (query-by-example) spoken term detection. The objective of this task is to find all lexical instances of a spoken term within an audio database of spoken words. Within the Mediaeval 2013 Spoken Web Search evaluation (which I helped organize), we have considered the scenario where nothing is known a-priori about the query or the database (only the fact that the database contains data from nine different languages was known).

Next, I discussed about how low/zero resource techniques can complement high resource systems. For this I hypothesized how babies learn about their surrounding world by registering similar repeating patterns that occur many times. I proposed the discovery of repeating information (e. g., repeating acoustic patterns in speech) to be used as an informed initialization to transcription systems to automatically retrieve sizable training corpora from high-quality seed transcriptions.

Finally, the following set of open questions were posed to the audience to trigger some discussion:

- How to incorporate acoustic modeling into dynamic programming techniques?
- How to describe the acoustic space (or whatever space) in an unsupervised (but robust) manner?
- How do we discriminate between “interesting/relevant” and “filler” events?

- Does it all make any sense? Or will there be a point where we always have enough training data for a given task?

3.2 Interpreting ‘Intentional’ Behaviour in Audio Scenes

Roger K. Moore (University of Sheffield, GB)

License © Creative Commons BY 3.0 Unported license
© Roger K. Moore

Whilst there is no doubt about the immense practical benefits that could be derived from the automated analysis of audio scenes, it is not clear that the research community has yet developed a sufficiently sophisticated theoretical framework to realise its full potential. Recent years have seen measurable progress in computational approaches to information extraction by applying the latest machine learning techniques to annotated (or even unannotated) data, but most of the focus has been on classifying the surface phenomena associated with acoustic events. Little attention has been given to interpreting the underlying ‘intentional’ states that are unique to living organisms and which drive the physical actions that are performed (particularly communicative behaviour such as speech). Of course, if there was a simple one-to-one relationship between internal intentional states and the consequent surface behaviour, then interpretation would be relatively straightforward. However, in reality there is significant ‘coupling’ (i. e., dependencies) between objects, agents and their environment, and this means that interpreting what is happening in an acoustic scene requires a yet-to-be-defined unified computational modelling approach which is capable of integrating the relevant contingencies. This stimulus talk illuminated these issues and raised the following issues for discussion: How important is it that we acknowledge that the world contains intentional agents? Can we envisage a unified computational modelling approach which is capable of integrating the relevant contingencies? What are the implications of modelling self (recursion, context dependency, embodiment)? And can an agent ever understand a natural scene if it is not (or has never been) part of it? The final question is thus: what does an automated agent need to know about the world and the entities it contains in order to make sense of a general audio scene?

3.3 Semantic-Affective Models for Multimodal Signal Processing

Alexandros Potamianos (National TU – Athens, GR)

License © Creative Commons BY 3.0 Unported license
© Alexandros Potamianos

In this stimulus talk, I reviewed multimodal aspects of audio processing focusing mainly on three areas. First, I considered the area of affective analysis and recognition of audio and multimedia streams. I presented recent results of emotion recognition from audio signals as well as movies, and the interplay between audio events, music and spoken language was outlined. Second, I discussed aspects of saliency and attention for audio and multimedia streams. Recent progress on selectional attention models for speech and audio were reviewed. Furthermore, the role of saliency/attention for audio/speech processing was discussed and future research directions outlined. Third, I addressed the topic of associative and representational models of audio and multimodal semantics. Motivated by cognitive considerations,

associative lexical semantic models have been recently proposed. These models have been extended to include also multimodal or crossmodal information such as images. I discussed how such models can be extended for audio, music, and speech content to create multimodal similarity networks as well as how such networks are relevant for inference and classification tasks.

3.4 Exploitation of Human Perception Principles in Audio Processing Systems

Gaël Richard (Telecom ParisTech, FR)

License  Creative Commons BY 3.0 Unported license
© Gaël Richard

The integration of auditory perception in most audio processing systems remains limited. A number of perceptually-relevant concepts have been exploited in audio processing research. But, for example for audio indexing/classification, it still seems that it is difficult to build a fully perceptually-relevant system that outperforms efficient machine-learning based methods that use only some rudimentary principles of perception. This remains surprising because from a pure acoustical point of view it intuitively appears that it may be unnecessary to capture similarity or dissimilarity information that is not perceived by humans. Should we look for better perceptual features or better perceptual representations such as cortical representations? Should we better model feature dynamics? Should we better model the complex and hierarchical processing information in the brain? In this stimulus talk, I discussed some of these issues. One may draw the conclusion that even though human perception principles are in general seen as important, mimicking the human perception or the functioning of the brain does not seem to be a prerequisite for computational audio analysis systems.

3.5 Stop Listening to Speech, Language, and Vision Research!

Paris Smaragdis (University of Illinois at Urbana Champaign / Adobe, US)

License  Creative Commons BY 3.0 Unported license
© Paris Smaragdis

Artificial intelligence, and more recently machine learning, has always been guided by the dream of making machines that can understand the world around them. Unfortunately the lion's share of this activity has been on domains that exhibit few insights on how to make machines that can listen, and this has resulted in an ongoing derailment of how machine learning should be applied on audio problems. A big chunk of machine learning is dealing with problems that stem from vision, language, and speech, which are all domains where we make hard decisions. A pixel either belongs to one object or another, a word is either “cat” or “dog”, and a spoken language belongs to one of many different families. As a result, the vast majority of machine learning approaches operate with a winner-takes-all philosophy, where the objective is to find that one solution that is the only correct value. Sound however is different. In real-life we never hear one sound or another. Instead, we hear mixtures of sounds. Such problems cannot be properly treated with tools such as discriminative learning and even many of the common generative models, and we see a need for some fundamental rethinking of how learning algorithms should be applied on sound.

3.6 Sound Event Detection and Recognition in Everyday Environments

Tuomas Virtanen (Tampere University of Technology, FI)

License © Creative Commons BY 3.0 Unported license
© Tuomas Virtanen

For humans, the most important functionality of auditory scene analysis is to acquire information about our everyday environments: a car approaching from behind, warning beeps, door knocking, door opening and closing and so on. Until now, most of the research on computational audio analysis has been done in the context of speech and music processing. Research on automatic detection and recognition of sound events has mostly been limited to isolated sound events, specific environments (such as meeting rooms), and small number of specific types of events.

Computational audio analysis in everyday environments has applications in areas such as multimedia content analysis, context-aware devices, assistive technologies, and acoustic monitoring. The research in the field has so far approached the task by studying two problems. The first one is context recognition, where a recording is classified into one of predefined contexts. Such contexts may be characterized by locations such as home, office, street, or grocery store or by physical and social activities. Second, sound event or acoustic event detection aims at estimating the start and end times of individual events in a recording as well as estimating a class label for each detected event.

Automatic detection and recognition of events in realistic environments requires addressing many problems (e. g., robustness) that have already been faced in the context of speech and music recognition. Additionally, operating with realistic sounds in everyday environments raises many new questions. Sound events can originate from very different sources and have therefore diverse acoustic characteristics, which may force us to rethink our conventional pattern recognition approaches. The identity of an event can be encoded in many different ways, e. g., by the rough shape of the spectrum, modulations over time, relationships between atomic sound units, or parts of a signal. This affects not only the feature extraction front-end used by an event detection system, but also the architecture of the whole system.

Finally, a single sound can have multiple different interpretations, and it is not at all clear how the event classes to be detected should be defined. Possibilities include definitions about the physical source, semantics, or acoustic similarity. Since it may not be possible to manually define classes for all sounds present in a signal, the use of unsupervised learning techniques needs to be taken into account.

4 Group Sessions

4.1 Small Data (Learning from Few Examples)

Bryan Pardo, Xavier Anguera, Jonathan Driedger, Bernd Edler, Jort Gemmeke, Franz Graf, Gernot Kubin, Frank Kurth, Meinard Müller, and Christian Uhle

License © Creative Commons BY 3.0 Unported license
© Bryan Pardo, Xavier Anguera, Jonathan Driedger, Bernd Edler, Jort Gemmeke, Franz Graf, Gernot Kubin, Frank Kurth, Meinard Müller, and Christian Uhle

We define “small data” in opposition to the current buzzword “big data”. These are cases where there are few labeled examples to learn from. This may be because the labeling requires intense analysis by an expert (e. g., structural analysis of a Beethoven symphony). This

may be because there are only a few examples in existence (there are only nine Beethoven symphonies), yet we wish to learn something useful and meaningful for some task. One concrete example of a “small data” situation is that of building an acoustic car crash detector for a particular tunnel. Collecting real data (crashing cars) is expensive. Car crash events are relatively rare in the tunnel (once or twice a year). Yet we want the detector to start working as soon as possible and with as few examples as possible. Another example is voice-controlled home automation for people with unique speech impairments. Here, again, data is hard to collect and each speaker is very different from the rest of the population.

There is an intellectual appeal to learning from small numbers of examples. Humans often learn generalizations from as few as three examples of a class. It would be interesting to learn how to duplicate this kind of performance. The practical appeal of being able to learn in an online manner, before collecting a lot of data (e.g. the car crash example) is also great. Further, when working with systems on small data sets, the researchers themselves can have a much better understanding of the data. Rather than millions of unexamined examples, there are dozens of well-understood ones. Current approaches to “small data” typically use statistical learners that require lots of data to work properly. Therefore approaches tend to look for ways to bridge the gap between learners that need lots of data and tasks that provide small numbers of examples. Approaches fall in the following categories:

- **Data Synthesis:** Create synthetic data by adding noise (of some expected variety) to the small number of known examples. Alternately synthesize data by using a simulation that can be done more cheaply than collecting real data (e.g., replace cars with garbage cans and crash garbage cans together in the tunnel).
- **General to Specific:** Start with a generic model, learned from a lot of data. Tweak that model slightly to conform to the particular “small data” case.
- **Model Selection:** Build several generic models for known generic cases, then collect a small number of data points from the current case to select which generic model best suits the current case. This can then be used in combination with the previous strategy to make it more specific.
- **Think Smarter:** Offload the learning to the human, who figures out a smarter way of preprocessing data to put it in a format that is very easy to learn from, by using extremely salient features.

The question we had was if there are other approaches we have not encountered. Can we do better than these four approaches?

4.2 Interaction and Affect

Martin Heckmann, Murtaza Bulut, Carlos Busso, Nick Campbell, Laurence Devillers, Anna Esposito, Sungbok Lee, Roger Moore, Mark Sandler, Khiet Truong, and Rita Singh

License © Creative Commons BY 3.0 Unported license
 © Martin Heckmann, Murtaza Bulut, Carlos Busso, Nick Campbell, Laurence Devillers, Anna Esposito, Sungbok Lee, Roger Moore, Mark Sandler, Khiet Truong, and Rita Singh

Interaction is driven by intentional agents. Agents accommodate to the role and capabilities of other agents. The success of the interaction depends on the generation and interpretation of appropriate signals, often across multiple modalities (e.g., bio-signals, image, speech). However, the effective processing of these signals also depends on “rich” information and not just “big” data. This includes the temporal and situational context as well as the role, characteristics, behaviors and traits of the agents. Current systems depend on theory-laden annotations (not capturing the true nature of the interaction), which unnecessarily constrain

the learning outcome. We believe that a viable first step is to develop and continuously adapt the model for the agent’s world from observed data. One possible means to achieve this is to implement algorithms for detecting changes and deviations from the learned normal/stable points.

4.3 Learning of Units and Knowledge Representation

Florian Metze, Xavier Anguera, Sebastian Ewert, Jort Gemmeke, Dorothea Kolossa, Emily Mower Provost, Björn Schuller, and Joan Serra

License © Creative Commons BY 3.0 Unported license
© Florian Metze, Xavier Anguera, Sebastian Ewert, Jort Gemmeke, Dorothea Kolossa, Emily Mower Provost, Björn Schuller, and Joan Serra

Our group came together to discuss how knowledge could be used to define and infer units of sound that could be used in a portable way for a number of tasks. Participants felt that a top-down approach would be needed, which is complementary to purely data-driven bottom-up clustering approaches, as are currently prevalent in classification experiments. Members wanted to specifically investigate how an attempt to solve multiple problems at the same time (“holistic” approach) could benefit each individual task by exposing and exploiting correlations and complementarity, which would otherwise stay hidden. Members also felt that a sound statistical framework was needed and that a careful modeling of uncertainty and a mechanism to feed back confidences was needed. This would also be beneficial in the presence of multiple, possibly overlapping signals as is typically the case for sounds. Finally, members were interested in working on meta-data of speech. First ideas were discussed on how to learn from data units representing emotions that would be both acoustically discriminative and useful in the context of a certain application, or discernible by humans.

Most members had some background in low-level feature extraction and in deep learning. Against this background, members developed an experiment, which they intend to execute in a distributed collaboration over the next couple of weeks. The experiment will be performed on the IEMOCAP database using various existing tools available to the group members. Collaboration tools will be set up at CMU. To establish a baseline, members will investigate the suitability of multi-task learning by training a single deep neural network (DNN) to predict both binary and continuous valued emotion targets on the IEMOCAP benchmark database. The network will be adapted to other databases (most likely AVEC and CreativeIT) to investigate the portability of the learner and to investigate the utility of multi-task learning. These experiments can be performed with feed-forward as well as recurrent architectures. Next, prior knowledge will be incorporated into the classification by adding database information, speaker information, or other meta-data (automatically extracted or manually labeled) as additional inputs to the network training. Finally, the recurrence loop will be optimized by investigating which information should be fed back. This information may comprise the utility of certain features or classes in a certain task, the saliency of some features, or the classification accuracy (posterior probabilities) of some classes on a held-out dataset. Members discussed an uncertainty weighted combination approach that should be able to update the structure and parameters of the classifier so as to improve classification accuracy. The goal will be to optimize the allocation of parameters towards modeling useful target units rather than attempting to accurately model distinctions that will eventually not be used in an application. Results will be published in peer-reviewed literature, and will hopefully lead to follow-up collaborations including organizing future workshops and joint proposals.

4.4 Unsupervised Learning for Audio

Tuomas Virtanen, Jon Barker, Shrikanth Narayanan, Alexandros Potamianos, Bhiksha Raj, Gaël Richard, Rita Singh, Paris Smaragdīs, Stefano Squartini, and Shiva Sundaram

License  Creative Commons BY 3.0 Unported license
© Tuomas Virtanen, Jon Barker, Shrikanth Narayanan, Alexandros Potamianos, Bhiksha Raj, Gaël Richard, Rita Singh, Paris Smaragdīs, Stefano Squartini, and Shiva Sundaram

After a more fundamental discussion on unsupervised learning for audio, our group decided to focus on the use of unsupervised learning in a concrete application scenario. Even though unsupervised learning could be used in many processing stages of a computational audio analysis system (e. g., to develop the feature extraction front-end), in practical scenarios one often takes advantage of some prior information. In particular, defining a specific application and ways to evaluate the performance of the audio processing system already imposes some prior information.

We considered an application where the goal is to detect car crash sounds from continuous audio recordings. In the unsupervised learning scenario, one has audio recordings that can be used as a training data, but no reference times of crashes are actually annotated. There are several kinds of prior information that can be used in the given application. First, one may assume that the events one is looking for can be characterizing by a specific set of audio features such as MFCCs. Second, one may assume to have a metric that is appropriate for describing distances between acoustic features. Third, we know that events are rare, i. e., only a small number of target events are present in the training data. Finally, we assume that each event is localized in time and the duration of events is approximately known (e. g. one or two seconds). The above prior information can be used for novelty-based audio segmentation using the calculated features and the distance metric. Alternatively, unsupervised learning can be used to learn features and a distance metric. The resulting segments can be assumed to be homogeneous. Segments can then be clustered so that each cluster contains a specific kind of sound. Subsequently, the developer of the system can manually examine each cluster to see whether a cluster contains a sound relevant for the development of the detector. Assuming that the events of interest are rare, the cluster with the largest number of segments need not be examined (containing sounds that do not correspond to a car crash). The system could also work in an incremental fashion, where the clustering may change as new data becomes available. This results in a system that achieves a more knowledgeable perspective on the problem to be solved.

The main benefits of the use of unsupervised learning in this application is the reduction of amount of manual work: the events of interest can be found from the recordings simply by examining a single sample from a cluster. In our discussions, it was also pointed out that the use of unsupervised learning removes a user bias and allows for finding phenomena or concepts that cannot be precisely defined.

4.5 Source Separation

Christian Uhle, Jonathan Driedger, Bernd Edler, Sebastian Ewert, Franz Graf, Gernot Kubin, Meinard Müller, Nobutaka Ono, Bryan Pardo, and Joan Serrà

License © Creative Commons BY 3.0 Unported license
© Christian Uhle, Jonathan Driedger, Bernd Edler, Sebastian Ewert, Franz Graf, Gernot Kubin, Meinard Müller, Nobutaka Ono, Bryan Pardo, and Joan Serrà

Our group attracted participants from various research areas to discuss aspects of source separation for audio signals, performed either in a blind way or by using additional knowledge about the underlying sources or the mixing process. In many source separation approaches, one assumes that sources are independent, uncorrelated, and do not overlap with regard to a given representation. Also one often presupposes that the mixing process is linear and time-invariant. However, in practice these assumptions are often violated. In addition, sound sources may influence or interact with each other, so that the separated source signals may sound unnatural or different to situation where they occur in an isolated fashion. Examples are the coupling between piano strings and the Lombard effect that describes the adaptation of a speaker to noisy environments. Further fundamental problems in source separation are the unmasking of undesired sounds (e. g., FM noise or audio coding artifacts), shortcomings of objective evaluation metrics, or the sound quality (e. g., due to the phase reconstruction problem). Last but not least, even the definition of what to understand by a source is ambiguous: a source can be a physical entity that emits sound, an object or event that is perceived by a human listener (stream), or a musical voice in a polyphonic sound mixture.

There are various applications that motivate ongoing research in source separation including remixing and upmixing, Karaoke applications, speech enhancement for hearing aids and communication, dialogue enhancement, audio editing, and audio content analysis. Besides these applications, source separation is a fascinating, intellectual, and interdisciplinary research area that requires and provides a deep understanding of the underlying audio material with regard to various aspects ranging from physical processes to cognitive aspects.

5 Further Topics

5.1 Engineering Selective Attention into Acoustic Scene Analysis Systems.

Jon Barker (University of Sheffield, GB)

License © Creative Commons BY 3.0 Unported license
© Jon Barker

A general goal of acoustic scene analysis is to recover abstract high-level descriptions of the individual sound sources given the raw acoustic mixtures. It is often assumed that a machine scene analysis system should extract some sort of ‘complete’ description in which all sources are described with equal detail. In certain contrived scenes, for example ‘cocktail parties’ composed of speakers uttering sentences from fixed grammars, computational systems are able to generate complete descriptions by composing individual source models and performing exact or approximate inference. In such cases machines can outperform human (e. g., [4]). However, it is unclear how such approaches can be usefully applied to handle complex everyday scenes containing unknown numbers of dynamically changing sources with unpredictable onsets and offsets.

In contrast to the above, ‘complete description’ problem, we can consider machine listening that adopts a more human version of scene analysis where there are favoured ‘attended sources’ (i. e., a ‘foreground’) and unattended sources that are allowed to remain unresolved in the background. Such systems would not form complete scene descriptions, but would instead try to mimic the human ability to fluidly switch attention between alternative ‘foregrounds’, driven by high-level goals or by the saliency of the competing sources (see [3]). A simple version of the approach is exemplified in the fragment decoding technique for robust speech recognition ([1, 2]): simple source-independent models are used to perform a local decomposition into acoustic ‘fragments’ and then, at a higher level, fragments are integrated over time by composing detailed models of the target speaker mixed with much simpler models of the background. However, within any attention-driven framework, where foreground and background are treated asymmetrically, there exist many unresolved questions. How can the complexity of the foreground and background models be balanced so as to maximise performance at a fixed computational cost? What are the dimensions of auditory salience that drive attention? How to model ‘top-down’ selective attention? How to model ‘bottom-up’ reflexive attention? In particular how much processing of the background is required in order to be aware of the salient qualities (particularly with respect to ‘top-down salience’)?

References

- 1 J. P. Barker, M. P. Cooke, and D. P. W. Ellis. *Decoding speech in the presence of other sources*. Speech Communication, vol. 45, no. 1, pp. 525, Jan. 2005.
- 2 N. Ma and J. Barker. *A fragment-decoding plus missing-data imputation system evaluated on the 2nd CHiME challenge*. Proceedings of the 2nd CHiME Workshop on Machine Listening in Multisource Environments, Vancouver, Canada, Jun. 2013, pp. 5358.
- 3 M. I. Posner and S. E. Petersen. *The attention system of the human brain*. Annual Review of Neuroscience 13: 2542, 1990.
- 4 S. Rennie, J. R. Hershey and P. A. Olsen. *Single Channel Multi-talker Speech Recognition: Graphical Modeling Approaches*. IEEE Signal Processing Magazine, Special issue on Graphical Models, November 2010.

5.2 Compensate Lexical/Speaker/Environment Variability for Speech Emotion Recognition

Carlos Busso (The University of Texas at Dallas, US)

License © Creative Commons BY 3.0 Unported license
© Carlos Busso

Affect recognition is a crucial requirement for future human machine interfaces to effectively respond to nonverbal behaviors of the user. Speech emotion recognition systems analyze acoustic features to deduce the speaker’s emotional state. However, human voice conveys a mixture of information including speaker, lexical, cultural, physiological and emotional traits. The presence of these communication aspects introduces variabilities that affect the performance of an emotion recognition system. Therefore, building robust emotional models requires careful considerations to compensate for the effect of these variabilities. Important research issues are concerned with normalization schemes that compensate the variability introduced by multiple communication aspects not related to emotions. These approaches include environment, speaker, and lexical normalization.

5.3 Interpretation and Computational Audio Analysis

Laurence Devillers (LIMSI – CNRS, FR)

License © Creative Commons BY 3.0 Unported license
 © Laurence Devillers
URL <http://www.limsi.fr/Scientifique/emotions/>

Most of the research on Computational Audio Analysis has been on classifying the surface phenomena associated with acoustic signals and with speech events. The meaning of these events usually depends on the context in which they occur. The analysis of audio (and video) scenes can help machines to interpret speech of humans or of human-machine interactions. One of the important issues is how to decide which contextual information to acquire and how to incorporate it into machine learning. Machines should be able to deal with interactions with multi-speakers and interpret the relationship between speakers. To give to the machines the capabilities to interpret and generate appropriate signals taking into account the context of the interaction (with multi-sources analysis) is a real challenge.

5.4 Perceptually Appealing Reconstruction of Spectrally Modified Signals

Jonathan Driedger, Meinard Müller (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
 © Jonathan Driedger and Meinard Müller

In many audio processing tasks such as source separation or time-scale modification, the audio signal is modified in the spectral domain and then resynthesized by applying some inverse transform. Examples are binary or relative masking in source separation procedures or phase propagation techniques as used in the phase vocoder. However, applying such modifications typically ignore the complex relationships between phases and magnitudes of superimposed sound components. As a result, besides the intended effects, the reconstructed signals often contain unwanted artifacts. In this seminar, we have raised the question of how to evaluate the quality of reconstructed signals. Further issues were how artifacts may be reduced using phase adaption strategies or perceptually masked using suitable post-processing techniques. A fundamental observation was that a listener's expectation of how a modified signal should sound often differs to what is actually contained in the data. This has shown that tasks such as time scale modification or source separation (without any further applications) are highly subjective and ill-posed problems.

5.5 The Situated Multimodal Facets of Human Communication

Anna Esposito (International Institute for Advanced Scientific Studies, IT)

License © Creative Commons BY 3.0 Unported license
 © Anna Esposito
Main reference M. Rojc, N. Campbell, (eds), "Coverbal Synchrony in Human-Machine Interaction," 434 pp., ISBN: 9781466598256, CRC Press, 2013.

Humans interact with each other through a gestalt of emotionally cognitive actions which involve much more than the speech production system. In particular, in human interaction,

the verbal and nonverbal communication modes seem to cooperate jointly in assigning semantic and pragmatic contents to the conveyed message by unraveling the participants cognitive and emotional states and allowing the exploitation of this information to tailor the interactional process. These multimodal signals consist of visual and audio information that singularly or combined may characterize relevant actions for collaborative learning, shared understanding, decision making and problem solving. This work will focus on the visual and audio information including contextual instances, hand gestures, body movements, facial expressions, and paralinguistic information such as speech pauses, all grouped under the name of nonverbal data, and on the role they are supposed to play, assisting humans in building meanings from them.

5.6 Bayes and Beyond Bayes: The Integration of Prior Knowledge

Sebastian Ewert (Queen Mary University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Sebastian Ewert

To analyse audio recordings using automated methods, one typically makes assumptions about characteristics and properties of the recorded content. Such assumptions can be explicit or implicit, and can exist on various semantic levels. For example, in music processing, methods often exploit that most musical instruments produce harmonic sounds to analyse the musical content or to identify individual sound sources. Similarly, in speech recognition, methods rely on the fact that different utterances of a specific phoneme are in a common manifold of a feature space, which can be described using probabilistic models. Even in methods, which are generally considered as unsupervised, one can find various implicit assumptions. For example, in methods such as NMF, one exploits that many sounds can be approximated by a convex combination of a few fixed spectral templates which would not be true for highly non-stationary sounds or noise. Also the number of templates used in NMF is typically based on some kind of assumption.

All these different assumptions can be considered as a form of prior knowledge and, in this sense, prior knowledge is an essential component in every signal analysis method. Still, it is not always clear how prior knowledge is integrated best. Some types of prior knowledge only loosely correlate with specific signal properties, and it might not be clear whether the integration of such prior knowledge is useful at all. It is also not always clear how the prior knowledge can be integrated. In particular, while prior distributions in Bayesian probabilistic models have been used successfully in recent years in this context, whether they can or should be used to represent a specific type of knowledge. Furthermore, prior knowledge can be available on various semantic levels. For example, a musical score provides high-level information about pitch and timing of note events, which can be used to simplify extremely complex problems such as source separation.

In this seminar, I asked and discussed with other participants the following questions. What kind of implicit and explicit prior knowledge are you facing in your work? How are you using prior knowledge in your methods? What kind of general strategies exist to integrate prior knowledge in front end transformations, in signal and acoustic models, in backend and machine learning components? What is your experience with prior knowledge on various semantic levels? What strategies do you employ to integrate knowledge beyond Bayesian modelling?

5.7 NMF meet Dynamics

Cédric Févotte (*JL Lagrange Laboratory – Nice, FR*)

License  Creative Commons BY 3.0 Unported license
© Cédric Févotte

Over the last ten years nonnegative matrix factorisation (NMF) has become a popular unsupervised dictionary learning and adaptive data decomposition technique with applications in many fields. In particular, much research about this topic has been driven by applications in audio, where NMF has been applied with success to automatic music transcription and single channel source source separation. In this setting, the nonnegative data is formed by the magnitude or power spectrogram of the sound signal and is decomposed as the product of a dictionary matrix containing elementary spectra representative of the data times an activation matrix which contains the expansion coefficients of the data frames in the dictionary.

In my own research, I have worked on model selection issues in the audio setting, pertaining to the choice of time-frequency representation (essentially, magnitude or power spectrogram), and to the measure of fit used for the computation of the factorisation. Driven by a probabilistic modelling approach, I came up with arguments in support of factorizing of the power spectrogram with the Itakura-Saito (IS) divergence [1]. Indeed, IS-NMF is shown to be connected to maximum likelihood estimation of variance parameters in a well-defined statistical model of superimposed Gaussian components and this model is in turn shown to be well-suited to audio. In my work, I have also addressed variants of IS-NMF, namely IS-NMF with temporal regularisation of the activation coefficients [2], automatic relevance determination for model order selection [3] and multichannel IS-NMF [4].

Recently, I have started to look into dynamical variants of NMF [5], in which structured transitions occur from spectral patterns to others. This is a desirable property for example for speech signals, for which some temporal correlation (or anti-correlation) is expected to occur between subset of speech patterns. Introducing dynamics into NMF is a challenging task at the modelling and estimation levels. To put it simply, one might say that NMF has superseded the traditional GMM. If HMM is the natural dynamical extension to GMM, what is the natural dynamical extension to NMF?

References

- 1 C. Févotte, N. Bertin, and J.-L. Durrieu. *Nonnegative matrix factorization with the Itakura-Saito divergence with application to music analysis*. *Neural Computation*, 21(3):793–830, Mar. 2009.
- 2 C. Févotte. *Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization*. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.
- 3 V. Y. F. Tan and C. Févotte. *Automatic relevance determination in nonnegative matrix factorization with the beta-divergence*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1592–1605, July 2013.
- 4 A. Ozerov and C. Févotte. *Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation*. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3):550–563, Mar. 2010.
- 5 C. Févotte, J. Le Roux, and J. R. Hershey. *Non-negative dynamical system with application to speech and audio*. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.

5.8 Features beyond Machine Learning

Martin Heckmann (Honda Research Europe, DE)

License  Creative Commons BY 3.0 Unported license
© Martin Heckmann

A large part of my research concentrates on the extraction of features from speech signals: on the one hand for recognizing what was said and on the other hand how it was said. In the development of these features I try to combine three ingredients: first, the usage of domain knowledge; second, taking inspirations from what is known of the processing in the human brain (e.g. high dimensional sparse representations); and, third, machine learning approaches. One example is a set of hierarchical spectro-temporal features, which build on a perceptual representation and form sparse and high-dimensional features learned from unlabeled speech data [1]. Currently, I am particularly interested in the extraction of more subtle prosodic variations which play a very important role in human communication. This includes back-channels which indicate how the listener is following the conversation as well as the prominence different words receive which is related to the importance a speaker attributes to a word [2, 3]. Here, the domain knowledge is one of the key ingredients so far. However, in recent years I experience a trend away from extensive domain knowledge and psychophysical inspirations more towards approaches based on machine learning. The different paralinguistic challenges at INTERSPEECH by Schuller et al. are a prime example as how the same set of features can successfully be applied to many different tasks with the right machine learning backend [4]. Related but a bit different is the tremendous success of Deep Neural Networks in the last two years. Currently they are used as a powerful and versatile tool of machine learning which is particularly suited to exploit the rich information provided by very large datasets. Furthermore, researchers have also started trying to integrate inspirations from the processing of the human brain in this approach such as convolutional networks. In this seminar, I have discussed ideas for methodologies to fruitfully integrate the rapid advances in machine learning with processing principles in the brain and domain knowledge to come up with better features.

References

- 1 M. Heckmann, X. Domont, F. Joublin, and C. Goerick. *A hierarchical framework for spectro-temporal feature extraction*. *Speech Communication*, Vol. 53, No. 5, pp. 736–752, 2011.
- 2 M. Heckmann. *Audio-visual evaluation and detection of word prominence in a human-machine interaction scenario*. In *Proc. INTERSPEECH*, Portland, OR, 2012.
- 3 M. Heckmann. *Inter-speaker variability in audio-visual classification of word prominence*. In *Proc. INTERSPEECH*, Lyon, France, 2013.
- 4 B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan. *Paralinguistics in speech and language—State-of-the-art and the challenge*. *Computer Speech & Language*, Vol. 27, No. 1, pp. 4–39, 2013.

5.9 Detection of Repeated Signal Components and Applications to Audio Analysis

Frank Kurth (Fraunhofer FKIE – Wachtberg, DE)

License © Creative Commons BY 3.0 Unported license

© Frank Kurth

Main reference F. Kurth, “The Shift-ACF: Detecting Multiply Repeated Signal Components,” in Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA’13), pp. 1–4, IEEE, 2013.

URL <http://dx.doi.org/10.1109/WASPAA.2013.6701820>

Our work in the last two years was mainly concerned with the detection of structured audio components within source signals. In this, an important type of structure are repetitions such as repeating bird calls or percussive elements in music. A few months ago, we have proposed a novel technique for detecting multiply (i. e., more than once) repeated signal components within a target signal. For such cases, we were able to improve classical autocorrelation techniques. In our experiments, we up to now have successfully considered applications in bioacoustics and in speech processing. It was interesting to discuss the topic within an interdisciplinary community as it was present at the Dagstuhl seminar and to learn about further possible applications—and existing solutions—from other domains, especially when dealing with noisy or distorted signals. For me, related interesting questions are both how to automatically separate, or even extract, all structured signal parts from the residual signal and how to do this efficiently for large scale signal scenarios.

As a first follow-up activity to the Dagstuhl seminar, I am organizing a special session on “Audio Signal Detection and Classification” covering topics such as audio monitoring, signal detection, segmentation and classification, audio fingerprinting, matching techniques, and audio information retrieval. The special session, which will be held at the IEEE Workshop on Cloud Computing for Signal Processing, Coding and Networking (IWCCSP) on March 11, 2014, aims at bringing together experts from the audio signal processing area with the cloud computing community.

5.10 Informed Source Separation for Music Signals

Meinard Müller, Jonathan Driedger (Universität Erlangen-Nürnberg, DE)

Joint work of Sebastian Ewert

License © Creative Commons BY 3.0 Unported license

© Meinard Müller and Jonathan Driedger

Main reference S. Ewert, B. Pardo, M. Müller, M. Plumbley, “Score-Informed Source Separation for Musical Audio Recordings,” IEEE Signal Processing Magazine, 2014, to appear.

One central problem in music processing is the decomposition of a given audio recording of polyphonic music into components that correspond to the various musical voices or instrument tracks. The main challenge arises from the fact that musical sources are highly correlated, share the same harmonies, follow the same rhythmic patterns, and so on. Musicians play together, follow the same lines, and interact with each other. As a consequence, different musical voices often do not differ statistically from each other, which makes the separation of musical sources or voices infeasible and even ill-defined problem. Therefore, when processing music data, music-specific techniques are needed that exploit musical knowledge or music-specific constraints. For example, to support the separation of musical voices, one strategy is to use additional cues such as the musical score or user input [1]. In this seminar,

we have discussed questions such as: Is source separation of music signals a meaningful problem? What is it good for? What are possible applications? How can one measure the success and the complexity of the task? How can one integrate additional knowledge? Where does one obtain such knowledge from? How can this knowledge be learned from example data? What can be learned from the field of speech processing?

References

- 1 S. Ewert, B. Pardo, M. Müller, M. Plumbley. *Score-Informed Source Separation for Musical Audio Recordings*. IEEE Signal Processing Magazine, 2014.

5.11 Approaching Cross-Audio Computer Audition

Björn Schuller (TU München, DE)

License © Creative Commons BY 3.0 Unported license
© Björn Schuller

Main reference B. Schuller, “Intelligent Audio Analysis,” 345 pp., Signals and Communication Technology Series, Springer, 2013.

URL <http://www.springer.com/engineering/signals/book/978-3-642-36805-9>

Substantial progress has been made over the last years in a number of intelligent audio analysis sub-disciplines that lead closer to the realisation of genuine cross-audio computer audition. This includes in particular advances in blind audio source separation such as by Non-negative Matrix Factorisation variants, but also in the feature extraction and computational intelligence parts, e. g., by feature brute-forcing, or context-sensitive deep neural networks and tandem architectures with graphical model topologies, or recent transfer learning approaches. By these and further means, the community is at a point where we are able to shift more into handling complex compounds of speech, music, and sound simultaneously as this is how they appear in the real world [1]. In this seminar, we have discussed important tools and inspirations on how to proceed on this avenue.

References

- 1 Björn Schuller. *Intelligent Audio Analysis*. Signals and Communication Technology, Springer, 2013.

5.12 What can we Learn from Massive Music Archives?

Joan Serrà (IIIA – CSIC – Barcelona, ES)

License © Creative Commons BY 3.0 Unported license
© Joan Serrà

Music is an extremely powerful means of communication that shapes our brain in intricate ways, unique to mankind, and transversal to all societies. As a scientific community we are slowly but steadily progressing towards the availability of massive amounts of music and music-related data for research purposes. The Million Song Dataset, Peachnote, the Yahoo! Music Dataset, the Last.fm API, Musicbrainz, or Wikipedia are just but some examples. Certainly, such big data availability will shift the perspective in which we approach many (if not all) of the traditional music information retrieval tasks. From genre or mood classification to audio or cover song identification, practically all tasks will experience a change of paradigm that frame them under more realistic, large-scale scenarios. In this seminar, we discussed new

avenues for research that are awaiting for us. In particular, future work will be concerned about extracting and using knowledge that can be distilled from such massive amounts of data—not only knowledge about music itself (rules, patterns, anti-patterns, and their evolution), but also knowledge about ourselves, as music listeners, users, or creators.

5.13 Acoustic Monitoring in Smart Home Environments: A Holistic Perspective

Stefano Squartini (Polytechnic University of Marche, IT)

License © Creative Commons BY 3.0 Unported license
 © Stefano Squartini
Joint work of Emanuele Principi
Main reference E. Principi, S. Squartini, F. Piazza, D. Fuselli, M. Bonifazi, “A distributed system for recognizing home automation commands and distress calls in the Italian language,” in Proc. of the 14th Annual Conf. of the Int’l Speech Communication Association (INTERSPEECH’13), pp. 2049–2053, ISCA, 2013.
URL http://www.isca-speech.org/archive/interspeech_2013/i13_2049.html

In recent years, there has been significant interest around the “Smart Home” paradigm, a scenario where several research fields seem to naturally converge. One of the most relevant objectives consists in monitoring the activity of inhabitants for different purposes: emergency state recognition and fall detection (especially for elderly people), intrusion or theft detection, people localization, usage of appliances, or power consumption besides the more common home automation commands, which have been already implemented in many commercial entertainment-oriented devices. In this context, acoustic monitoring techniques play an important role. Even though many scientific studies have been conducted so far, the results do not yet seem to match the market expectations.

Our research group is developing distributed system for recognizing home automation commands and distress calls in Italian language. The system integrates the automatic recognition of emergency states and home automation commands with remote assistance and hands-free communication. The ITAAL database has been developed for this purpose and a preliminary prototype is already available. Nevertheless, many issues still need to be addressed in order to make the system more appealing, reliable and useful for exploitation in real domestic environments. This typically requires dealing with heterogeneous acoustic data, which must be treated by looking at them from a holistic perspective, also taking other types of sensing activity into account. Some of these issues are reported here as open challenges to be addressed in future research:

- How to integrate speech and sound analysis for activity monitoring? Utterances spoken by a user, even if not really related to specific commands devoted to activate certain smart functionalities, can be useful to understand what the user is doing and in which part of the house he is located, specially if adequately integrated with no-speech sounds related to his activity.
- How to integrate information coming from infra- and ultra-sound sensors? Spanning the frequency range beyond the audible range can be very useful (e. g., subsonic sounds for fall detection and ultrasonic sounds for localization), especially in an integrated fashion with the “real” acoustic information. Therefore, unsupervised learning techniques can be implemented to find out and efficiently use cross-domain relationships.
- What is the role of paralinguistic features? In emergency state recognition, for instance, the capability of detecting the presence of paralinguistic features in the vocal activity,

and likely understanding their meaning, can have a substantial impact in the overall performance and asks for consideration in smart home environments.

- How to deal with minimum a-priori knowledge? In several practical smart home scenarios, the adaptation of automatic recognition systems to a speaker’s characteristics is not allowed, since the provided technology should be as transparent as possible to the final user.
- How to deal with the “novelty” issue? One of the objectives of acoustic monitoring consists in automatically recognizing a novel event with respect to the “usual” ones, in order to take adequate actions (e. g. in case of thefts).

5.14 Sound Processing in Everyday Environments

Emmanuel Vincent (INRIA – Nancy – Grand Est, FR)

License  Creative Commons BY 3.0 Unported license
© Emmanuel Vincent

I am interested in an efficient integrated approach to sound processing in everyday environments. The various relevant tasks are often treated one after another: source localization, source separation, speaker/event identification, speech recognition. This “pipeline” approach yields suboptimal results due to the propagation of errors from one step to the next. Our approach is to propagate not only deterministic signals and values but a full posterior distribution (which is approximated as a Gaussian) from one step to the next. Some techniques exist to estimate this distribution but they are not very accurate yet. Burning questions in this context are: How to accurately estimate and propagate uncertainty? How to use it in combination with state-of-the-art ASR and speaker/event identification systems?

References

- 1 N. Duong, E. Vincent, and R. Gribonval. *Spatial location priors for Gaussian model based reverberant audio source separation*. EURASIP Journal on Advances in Signal Processing, Springer, 2013.
- 2 A. Ozerov, M. Lagrange, and E. Vincent. *Uncertainty-based learning of acoustic models from noisy data*. Computer Speech and Language, Elsevier, 2013, 27 (3), pp. 874-894.
- 3 J. Barker, E. Vincent, N. Ma, H. Christensen, and P. Green. *The PASCAL CHiME Speech Separation and Recognition Challenge*. Computer Speech and Language, Elsevier, 2013, 27 (3), pp. 621-633.
- 4 E. Vincent, J. Barker, S. Watanabe, J. Le Roux, F. Nesta, and M. Matassoni. *The Second ‘CHiME’ Speech Separation and Recognition Challenge: An overview of challenge systems and outcomes*. IEEE Automatic Speech Recognition and Understanding Workshop, 2013.

6 Towards a Definition of Computational Audio Analysis (CAA)

Towards the end of the seminar, each participant was asked to give some kind of definition for a research field we coined “Computational Audio Analysis” (CAA). The following list gives an overview of the various statements which, as a whole, also give a good impression about the range of topics we have discussed at our Dagstuhl seminar.

- Computational audio analysis provides quantitative methodologies that enable detailed analyses of human behavior and interaction.

- The goal of computational audio analysis (CAA) is to understand the underlying structure of a given audio recording using computational methods in order to extract information and higher level semantics. CAA is a highly multidisciplinary field bringing together researchers in computer science, digital signal processing, machine learning with domain experts in, for example, speech and music processing, human computer interaction, biology, medicine, and acoustics.
- In the field of artificial intelligence (AI), people have tried to realize machines that simulate the role of a human. Examples are the Deep Blue system in chess or the Watson system for answering questions posed in natural language. Computational audio analysis aims at realizing machines that hear and understand sounds like a human. However, it seems to be a long way to realize such system, so that CAA remains an exciting research area.
- Computational audio analysis aims at inferring meaningful structures from audio signals, finding hidden relationships in heterogeneous collections of acoustic data from multiple perspectives, as well as detecting and understanding the meaning of events as occurring in natural environments
- The objective of computational audio analysis is to give a generative explanation of a sound complex, where a soundscape is decomposed with sufficient fidelity to meet the needs of particular applications.
- Computational audio analysis aims at extracting information from audio signals using techniques from signal processing, machine learning, information retrieval, and related fields. One central objective is to segment, structure, and decompose audio signals into elementary units that have some semantic (e. g. linguistic, musical) meaning. These units not only serve as basis for higher level analysis and classification tasks but also deepen the understanding of the underlying acoustic material.
- Computational audio analysis refers to the modeling and analysis of audio, in particular the voice, with the goal to extract ‘meaningful information’ from audio. What ‘meaningful information’ means, depends on the respective application. Inferring interactive events or states from audio, classifying environmental sound events, or separating sources to improve ASR are such examples tasks. Furthermore, CAA provides us with techniques to automatically attribute labels or perceptual characteristics to sounds.
- Computational audio analysis—in an utmost compact description—essentially focuses on extracting information from audio using computational methods.
- Computational audio analysis is the processing of audio signals in order to characterize or decode them in a way humans can understand. It incorporates signal processing techniques as well as models of perception and cognition. The main difficulty is that it needs to model a highly complex system with large inter-subject variability: human listeners.
- Computational audio analysis means resolving audio into machine understandable constructs.
- In computational audio analysis, annotation-flexible models that adapt to new conditions are developed in order to achieve a more representative (machine) learning outcome. Furthermore, the interplay between speech signals, other human-produced signals such as physiological signals (heart rate, skin conductance, activity, gestures), and non-speech audio signals (e.g. cough, snoring, sneezing) are explored. The understanding of how and why audio influences the human mind using low-level features can open possibilities for new application.

- Computational audio analysis means extracting knowledge from audio and making sense of it.
- Computational audio analysis provides computational methods for finding relevant structure (pertinent features and class labels as well as appropriate decompositions) in acoustic data, where relevancy, pertinence and appropriateness are usually defined in a task-dependent way. Methods are not limited to acoustic data, but can use multi-modal input, as long as audio data is among the considered modalities.
- Computational audio analysis aims at the detection, separation and description of acoustic objects via computational means. Descriptions can be of a qualitative (e. g., “warm”) and quantitative (e. g. “70 dB SPL”) nature. The source of these acoustic objects can be a human speaking, real or virtual musical instruments being played, or other vibrating physical objects such as loudspeakers. In addition to the analysis of separable acoustic objects, CAA also targets at holistic descriptions of acoustic scenes or parts of a scene (e. g., being at a train station).
- Computational audio analysis is the automated analysis of acoustic signals (whether natural or man-made) in order to perform some task that has utility to humans. There are no restrictions on the task: the setting may be online/offline, unimodal/multimodal, passive/interactive and may involve any form of acoustic signal including speech, music or environmental audio. The analysis may use perceptually motivated features (e. g., MFCCs) or perceptually motivated processing. However, in contrast to computational audition, the processing does not need to follow human audio processing, i. e., it does not explicitly model human hearing and the field is not concerned with learning about human hearing from human/machine comparisons.
- Computational audio analysis is a way to describe the effect that audio (both naturally occurring and artificially synthesized) has on humans, independent of language, linguistics, or phonetics. Due to the difficulty of describing its “targets” with words, or measure its physiological effects exactly, labels are very hard to get by. This makes CAA a challenging combination of fields such as computer science, musicology, psychology, or physiology.
- Computational audio analysis is the analysis and interpretation of an acoustic scene by a machine. This analysis can be either obtained in a supervised way, which is guided by human perception or sound production mechanism when known, or it can be unsupervised with the aim, for example, to discover new concepts (such as sound objects or sound primitives) not necessarily formally defined in advance by humans.
- Computational audio analysis is about machine-assisted extraction of information from sound. It can be either fully automatic (unsupervised) or user-guided (semi-supervised).
- Speech conveys information beyond verbal message including intentions, emotions, and personality traits that influence the way we communicate with others (people, robots, computers, devices). Computational audio analysis offers the opportunity to develop tools for learning and inferring these traits. The challenge in building such systems is to capture the temporal dynamic and situational context of behaviors.
- Computational audio analysis is the processing and modeling of the inherently heterogeneous general audio signal to uncover latent structures, to derive mappings to and between representations of interest, and to empower target applications such as summarization, retrieval, synthesis, and categorization. As a special case, the computational representations and formalism of CAA can benefit from human audition principles.
- Computational audio analysis is about processing audio data with respect to a specific application scenario and domain knowledge in order to extract task-specific information.
- Computational audio analysis is concerned with the extraction of a parametric description

- for an audio signal from its waveform (and possibly other additional representations). The type of the description varies depending on the requirements of the desired application.
- Computational audio analysis aims at understanding audio by means of computational means. This could mean being able to build a model of the source (source modeling and separation) extracting relevant messages (speech recognition), or understanding the environment the sources are in (e. g., room ID through reverb). CAA is open to any and all computational methods to do so (including semantic web, crowd sourcing).
 - Computational audio analysis involves the processing of audio signals by the help of computers with the objective to obtain information from it. Such information can refer different levels of abstraction ranging from basic signal measurements and low-level features to semantically meaningful information such as words, emotions, or melodies.
 - Computational audio analysis is the intersection of audio analysis by digital means (i. e., digital signal processing) with computer science. It therefore might include any relevant aspect of computer science, including but not limited to logic, inference, representation (ontologies), HCI, information retrieval, machine learning, cryptography and encryption, autonomous agents, communication (not telecommunication) theory, and so on. It should develop computational means and mechanisms for transitioning from audio data, to audio information, to (audio) knowledge and understanding for all forms of audio, i. e., speech, music, environmental, making that information and knowledge useable in a wide variety of application domains, including creative activity. It does not exist in isolation and has close ties to other sensory and affective data/modalities. It embraces the representational power of Semantic Web technologies which empowers many of the areas of computer science above in the linked data world of the future.
 - By audio, we deal with mechanical waves, i. e., a complex series of changes in or oscillation of pressure as compound of frequencies within the acoustic range available to humans and at sufficiently intense level to be perceived, i. e., audible by them. The analysis of audio aims at the extraction of information and, on a higher level, attachment of semantic meaning to audio signals. Computational audio analysis typically includes the involvement of computational intelligence algorithms as provided by the means and methods of machine learning going beyond signal processing.
 - Computational audio analysis deals with rich (audio) data and a complex (audio) signals. It encompasses a variety of aspects such as the analysis of spoken language, the mood of a song, and the human interaction including feelings and emotion.
 - Computational audio analysis is the engineering approach to reproduce the human capability of processing sounds to understand acoustic scenes and respond appropriately to the environment.
 - Computational audio analysis deals with the analysis of audio in combination with other sensor information such as video, body sensors, GPS, and so on. The analysis of such data is generally statistical, deep, atheoretical, and hard for people to understand. CAA should be time- and context-dependent. It may involve continuous adaptation and may incorporate protension.
 - Computational audio analysis is the use of computers (from microprocessors over smart-phones to supercomputers) for the analysis of audio signals (acquisition and storage, feature extraction, model building and interpretation) with applications in telecommuni-cations, multimedia, automotive, industry, biomedicine, performing arts, forensics, human curiosity, and science.

Participants

- Xavier Anguera
Telefónica Research –
Barcelona, ES
- Jon Barker
University of Sheffield, GB
- Stephan Baumann
DFKI – Kaiserslautern, DE
- Murtaza Bulut
Philips Research Lab. –
Eindhoven, NL
- Carlos Busso
The University of Texas at
Dallas, US
- Nick Campbell
Trinity College Dublin, IE
- Laurence Devillers
LIMSI – CNRS, FR
- Jonathan Driedger
Univ. Erlangen-Nürnberg, DE
- Bernd Edler
Univ. Erlangen-Nürnberg, DE
- Anna Esposito
Intern. Institute for Advanced
Scientific Studies, IT
- Sebastian Ewert
Queen Mary University of
London, GB
- Cédric Févotte
JL Lagrange Laboratory –
Nice, FR
- Jort Gemmeke
KU Leuven, BE
- Franz Graf
Joanneum Research – Graz, AT
- Martin Heckmann
Honda Research Europe, DE
- Dorothea Kolossa
Ruhr-Universität Bochum, DE
- Gernot Kubin
TU Graz, AT
- Frank Kurth
Fraunhofer FKIE –
Wachtberg, DE
- Sungbok Lee
Univ. of Southern California, US
- Florian Metze
Carnegie Mellon University, US
- Roger K. Moore
University of Sheffield, GB
- Emily Mower Provost
University of Michigan –
Ann Arbor, US
- Meinard Müller
Univ. Erlangen-Nürnberg, DE
- Shrikanth S. Narayanan
Univ. of Southern California, US
- Nobutaka Ono
National Institute of Informatics –
Tokyo, JP
- Bryan Pardo
Northwestern University –
Evanston, US
- Alexandros Potamianos
National TU – Athens, GR
- Bhiksha Raj
Carnegie Mellon University, US
- Gaël Richard
Telecom ParisTech, FR
- Mark Sandler
Queen Mary University of
London, GB
- Björn Schuller
TU München, DE
- Joan Serra
IIIA – CSIC – Barcelona, ES
- Rita Singh
Carnegie Mellon University, US
- Paris Smaragdis
University of Illinois at Urbana
Champaign / Adobe, US
- Stefano Squartini
Polytechnic Univ. of Marche, IT
- Shiva Sundaram
Audyssey Laboratories, US
- Khiet Truong
University of Twente, NL
- Christian Uhle
Fraunhofer IIS – Erlangen, DE
- Emmanuel Vincent
INRIA – Nancy – Grand Est, FR
- Tuomas Virtanen
Tampere University of
Technology, FI



Proxemics in Human-Computer Interaction

Edited by

Saul Greenberg¹, Kasper Hornbæk², Aaron Quigley³,
Harald Reiterer⁴, and Roman Rädle⁴

1 University of Calgary, CA, saul.greenberg@ucalgary.ca

2 University of Copenhagen, DK, kash@diku.dk

3 University of St Andrews, GB, aquigley@st-andrews.ac.uk

4 Universität Konstanz, DE, [Harald.Reiterer,Roman.Raedle]@uni-konstanz.de

Abstract

In 1966, anthropologist Edward Hall coined the term “proxemics.” Proxemics is an area of study that identifies the culturally dependent ways in which people use interpersonal distance to understand and mediate their interactions with others [1]. Recent research has demonstrated the use of proxemics in human-computer interaction (HCI) for supporting users’ explicit and implicit interactions in a range of uses, including remote office collaboration, home entertainment, and games. One promise of proxemics is the realization of context-aware environments, which have been extensively pursued since Mark Weiser’s seminal paper, “The computer for the 21st century,” written in 1991. However, the potential of proxemics in HCI is still underexplored and many research questions remain unanswered.

With the growing interest in using proxemics, we organized the Dagstuhl Seminar 13452 on the topic. “Proxemics in Human-Computer Interaction,” was held from November 3–8, 2013, and it brought together established experts and young researchers from fields particularly relevant to Proxemic Interactions, including computer science, social science, cognitive science, and design. Through an open keynote, mini talks, brainstorming, and discussion in breakout sessions, seminar attendees identified and discussed challenges and developed directions for future research of proxemics in HCI.

Seminar 03.–08. November, 2013 – www.dagstuhl.de/13452

1998 ACM Subject Classification H.5.2 User Interfaces: Theory and methods.

Keywords and phrases Proxemics, Proxemic Interactions, theory, vision, technology, application, distance, orientation, location, identity, movement, dark patterns

Digital Object Identifier 10.4230/DagRep.3.11.29

1 Executive Summary

Saul Greenberg

Kasper Hornbæk

Aaron Quigley

Harald Reiterer

Roman Rädle

License © Creative Commons BY 3.0 Unported license

© Saul Greenberg, Kasper Hornbæk, Aaron Quigley, Harald Reiterer, and Roman Rädle

Introduction

Over time, people encounter different dimensions of proxemics in everyday life, such as in face-to-face communication while discussing ongoing work with colleagues, in an elevator with strangers as private space is suspended, or at home with their families. In disciplines



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Proxemics in Human-Computer Interaction, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 29–57

Editors: Saul Greenberg, Kasper Hornbæk, Aaron Quigley, Harald Reiterer, Roman Rädle



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

like architecture and interior design, knowledge about proxemics has been used for decades to model use of space for face-to-face interactions, urban planning, and environmental design. In human-computer interaction (HCI) and human-robot interaction (HRI), the use of proxemics is fairly new, and both disciplines recently began employing proxemics and related theories and models (e.g., Hall's theory of proxemics in his book, "The Hidden Dimension" [2]) to design new interaction concepts that act on proxemics features. Several recent designs explore the use of human body position, orientation, and movement for implicit interaction with large displays, supporting collaboration, and to control and communicate with robots. This research is facilitated by the operationalization of proxemics for ubiquitous computing [16], toolkits to track proxemics [7, 8, 9], and new paradigms such as reality-based interaction (RBI) [4] or Blended Interaction [6] that take a fresh look at the role of the user's body and the environment in HCI. However, work on understanding how proxemics can be used for HCI (and HRI) has only just begun (e.g., Proxemic Interactions [1]).

Goals and Structure

In the seminar, we used Greenberg et al.'s dimensions on Proxemic Interactions [1] and Pedersen et al.'s Egocentric Interaction Paradigm [11] as starting points. These theories are based on findings regarding how humans perceive proxemics; therefore, they might be incomplete, particularly since human perception is much more subtle, gradual, and less discrete than illustrated in Hall's reaction bubbles (proxemic zones [2]). In addition, these discrete zones cope with only the physical features (perception of interpersonal distance). Other features, such as psychological and psychophysical features, have not yet been considered in HCI. However, these features are perceptible by human sensors (olfaction, equilibrioception, and thermoception). Current theories neither give guidelines nor provide sufficient methods for "good" or "bad" designs for systems employing proxemics.

We thought the time was right for bringing researchers with different backgrounds and experiences together to map out the important questions that remain unanswered and to generate ideas for developing an agenda for future research on proxemics in HCI.

The structure of the seminar was based on the four pillars technology, application, vision, and theory that were equally exposed in seminar activities. The forum held 29 attendees with multidisciplinary backgrounds from research institutes in Canada, Denmark, England, Switzerland, Australia, France, Belgium, and Germany. We achieved productive and critical reflections and prospects on proxemics in HCI by letting experts from their respective fields work on a shared vision and theory. We selected the attendees to ensure an equal distribution of expertise across the four pillars.

The diversified program allowed attendees to introduce themselves and their work in brief presentations and offered one impulse keynote given by Saul Greenberg and Nicolai Marquardt. Greenberg and Marquardt coined the term Proxemic Interactions and decisively influenced the application of proxemics in HCI. We also provided ample time for discussions, breakout sessions, and creative work addressing concepts such as:

- Intelligibility of Proxemic Interactions
- Users' options to opt-in or opt-out
- The "dark side" of Proxemic Interactions
- The meaning of physical space
- How image schemas [3] can be used to brainstorm innovative proxemic systems
- Ad-Hoc proxemics
- Including everyday entities in proxemic systems

Throughout the entire seminar, attendees were encouraged to write down their questions, ideas, and comments. These materials were collected and posted to one of the four pillars on a pin board for the purpose of inspiring breakout groups and ad lib collaboration. The breakout session proposed by the group centered around open problems and challenges within proxemic interactions, which was then discussed in each session.

Technology

In recent years, emerging technology has changed the interaction between human and computer. For instance, smartphones and tablets have entered our daily life. More of such novel post-WIMP¹ technologies will be available in the foreseeable future and ultimately define how we interact in physical spaces. Interaction might take place across device boundaries on (multiple) public [15], large and private, mobile, and tangible displays [13]. It might involve collaboration of co-located users around interactive tabletops [7], in front of large vertical screens [5], or on rollout displays [14]. It might be based on non-traditional, post-WIMP interaction styles, such as pen-based [10], multi-touch, and tangible user interfaces. Or, it might provide new forms of functionality beyond the traditional WIMP model of applications by tracking users' spatial location and movements for navigation within large, digital information spaces [12]. Attendees discussed existing technologies that allow people-to-people, people-to-object, and object-to-object proxemics relations tracking, as well as improvements on tracking reliability using sensor fusion.

Application

Seminar attendees discussed the “light” and “dark” side of Proxemic Interactions. Until now, research has focused on the benefits of these interactions; however, they bear risks. We all can imagine how advertisement would change if it becomes possible to show customized ads according to our online shopping profiles while we are walking on public streets or in shopping malls. During the seminar, participants discussed what types of applications would best showcase the benefit of proxemics and avoid the risks that arise when systems are able to track and identify people. Part of this discussion included brainstorming opt-in or opt-out functions for proxemics-aware systems so that users can remain in control of these systems.

Vision

In its past, HCI has benefited from ambitious visions of future interaction such as Apple's Knowledge Navigator or Mark Weiser's “A day in the life of Sal” [16]. Although visions are not always helpful and can lead in wrong directions, we believe that a new overarching vision of future Proxemic Interactions can help inspire ongoing research and thrive in coming generations. This vision is intended to inform researchers, designers, and laymen alike. For researchers, a vision can serve to illustrate research goals, trigger new research directions, and create awareness for as yet un-reflected assumptions in our field. For designers, visions help to present concepts and technologies as a part of a believable scenario – and not only in the isolation of conference papers. Furthermore, visions serve to fascinate and inspire laymen, who prefer to learn about future technologies from narrations instead of purely technical publications. The seminar aimed at creating a unified vision of Proxemic Interactions based

¹ WIMP stands as an acronym for Windows, Icons, Menus, Pointers

on the individual contributions and experiences of the seminar attendees. Current and past visions have been discussed in plenum and breakout groups.

Theory

In the light of the countless variants and dynamics of post-WIMP interaction, traditional collections of design guidelines or “golden rules” cannot provide enough guidance about “good” or “bad” designs. Instead, we need better theories and models of human cognition to be able to understand and classify designs of Proxemic Interactions and to predict their appropriateness. We wanted to understand how physical, psychological and psychophysical features collate and can be transferred into a coherent theory of proxemics in HCI and how to give guidelines or provide sufficient methods for “good” or “bad” designs. Therefore, we had to:

1. Better understand proxemics in HCI to develop such methods
2. Discuss the open question: to what extent can proxemics leverage or constrain human-computer interaction?

Conclusion

The Dagstuhl Seminar 13452 offered a fantastic forum for established researchers and practitioners at a comfortable place. We framed and discussed research questions and worked together on a unifying theory for Proxemics in Human-Computer Interaction. Applications for Proxemic Interactions were sketched out and critically reflected in the light of the “dark side” of proxemics. We also discussed how we can learn from related fields and how they can profit from proxemics in HCI.

The seminar can be seen as a good starting point to identify the role of Proxemics in Human-Computer Interaction. However, it still remains an open research area and its place in HCI needs to be better understood.

References

- 1 Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R. and Wang, M. 2011. Proxemic Interactions: The New Ubicomp? *interactions*. 18, January (2011), 42–50.
- 2 Hall, Edward, T. 1966. *The Hidden Dimension*. Doubleday.
- 3 Hurtienne, J., Israel, J.H. and Weber, K. 2008. Cooking up real world business applications combining physicality, digitality, and image schemas. *In Proc. of TEI'08*, New York, New York, USA, Feb. 2008, 239.
- 4 Jacob, R.J.K., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T. and Zigelbaum, J. 2008. Reality-based interaction: a framework for post-WIMP interfaces. *In Proc. of CHI'08*, New York, NY, USA, 2008, 201–210.
- 5 Jakobsen, M. and Hornbæk, K. 2012. Proximity and physical navigation in collaborative work with a multi-touch wall-display. *In Proc. of CHI EA'12*, New York, NY, USA, 2012, 2519–2524.
- 6 Jetter, H.-C., Reiterer, H. and Geyer, F. 2013. Blended Interaction: understanding natural human-computer interaction in post-WIMP interactive spaces. *Personal and Ubiquitous Computing*. (Oct. 2013).
- 7 Klinkhammer, D., Nitsche, M., Specht, M. and Reiterer, H. 2011. Adaptive personal territories for co-located tabletop interaction in a museum setting. *In Proc. of ITS'11*, New York, NY, USA, Nov. 2011, 107.

- 8 Marquardt, N., Diaz-Marino, R., Boring, S. and Greenberg, S. 2011. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. *In Proc. of UIST '11*, New York, NY, USA, 2011, 315–326.
- 9 Marquardt, N., Hinckley, K. and Greenberg, S. 2012. Cross-device interaction via micro-mobility and f-formations. *In Proc. of UIST'12*, New York, NY, USA, Oct. 2012, 13.
- 10 Matulic, F. and Norrie, M.C. 2013. Pen and touch gestural environment for document editing on interactive tabletops. *In Proc. of ITS'13*, New York, NY, USA, Oct. 2013, 41–50.
- 11 Pederson, T. 2012. Proximity as Key Property in the Egocentric Interaction Paradigm. http://hci.uni-konstanz.de/proxemics/wp-content/uploads/2012/10/proxemics2012_Pedersen_et_al.pdf.
- 12 Rädle, R., Jetter, H.-C., Butscher, S. and Reiterer, H. 2013. The effect of egocentric body movements on users' navigation performance and spatial memory in zoomable user interfaces. *In Proc. of ITS'13*, New York, NY, USA, 2013, 23–32.
- 13 Spindler, M., Stellmach, S. and Dachsel, R. 2009. PaperLens: Advanced Magic Lens Interaction Above the Tabletop. *In Proc. of ITS'09*, New York, NY, USA, Nov. 2009, 69–76.
- 14 Steimle, J. and Olberding, S. 2012. When mobile phones expand into handheld tabletops. *In Proc. of CHI EA'12*, New York, NY, USA, May 2012, 271–280.
- 15 Vogel, D. and Balakrishnan, R. 2004. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. *In Proc. of UIST'04*, New York, NY, USA, 2004, 137–146.
- 16 Weiser, M. 1999. The computer for the 21st century. *ACM SIGMOBILE Mobile Computing and Communications Review*. 3, 3 (1999), 3–11.

2 Table of Contents

Executive Summary

Saul Greenberg, Kasper Hornbæk, Aaron Quigley, Harald Reiterer, and Roman Rädle 29

Open Keynote

Proxemic Interactions

Saul Greenberg and Nicolai Marquardt 36

Overview of Talks

Investigating the Influence of Culture on Proxemic Behavior for Humanoid Robots
Elisabeth André 36

Virtual Proxemics

Jakob E. Bardram 37

Social Interaction in Palliative Care

Susanne Boll 38

Tangible Views into Rich Information Spaces in Proximity of Large Displays

Raimund Dachsel 38

The Meaning of Space in Interaction

Joern Hurtiénne 39

Hybrid-Image Visualization – or Perception-Based Proxemic Interaction

Petra Isenberg 39

Information Visualization and Proxemics: Design Opportunities and Empirical Findings

Mikkel R. Jakobsen 40

Proxemics for ad-hoc communities of devices

Hans-Christian Jetter 40

Technical Challenges of Sensing People’s and Devices’ Proxemic Relationships

Nicolai Marquardt 41

Proxemics as Play Resource

Florian Floyd Mueller 41

Situative Space Model – for human-centric ad-hoc smart environments

Thomas Pederson 41

User-Aware Devices: How Do we Gracefully Manage Imperfect Automation?

Stacey D. Scott 42

Resizable Mobile Devices for Ad-Hoc Mobile Meetings

Jürgen Steimle 42

Opportunities for Intelligibility in Proxemic Interactions

Jo Vermeulen 43

Siftor: subtle interaction in an art gallery context

Daniel Vogel 43

Working Groups

Play and Proxemics <i>Florian Floyd Mueller</i>	44
Intelligibility for Proxemic Interactions <i>Jo Vermeulen</i>	45
Dark Patterns in Proxemic Interactions <i>Saul Greenberg</i>	49
Ad-Hoc Proxemics – Inclusion of everyday entities in proxemics systems <i>Thomas Pederson</i>	53
Challenges of Sensing People’s and Devices’ Proxemic Relationships <i>Nicolai Marquardt</i>	55
Participants	57

3 Open Keynote

3.1 Proxemic Interactions

Saul Greenberg (University of Calgary, CA) and Nicolai Marquardt (University College London, UK)

License  Creative Commons BY 3.0 Unported license
© Saul Greenberg and Nicolai Marquardt

In the everyday world, much of what we do as social beings is dictated by how we interpret spatial relationships. This is called proxemics. What is surprising is how little people's expectations of spatial relationships are used in interaction design, i.e., in terms of mediating people's interactions with surrounding digital devices such as digital surfaces, mobile phones, tablets, and computers. Our interest is in proxemic interaction, which imagines a world of devices that have fine-grained knowledge of nearby people and other devices – how they move into range, their precise distance, their identity and even their orientation – and how such knowledge can be exploited to design interaction techniques. Just as people expect increasing engagement and intimacy as they approach others, so should they naturally expect increasing connectivity and interaction possibilities as they bring themselves and their devices in close proximity to one another and to other things in their everyday ecology. The joint introductory seminar by Greenberg and Marquardt introduced the notion of proxemics. It begins by stepping through a brief history of the evolution of HCI from user-centered design to present-day embodied interaction. It then introduces Hall's social science theory of proxemics, followed by variations of how others have developed this theory of proxemics to both refine and extend what is covered by it. The seminar then turned to proxemic interactions, which applies the theory to system design. It described how proxemics can be operationalized by what can be sensed and stored by computer, and then how a toolkit – the Proximity Toolkit – can simplify how programmers access and use this sensed data to build prototypes. A variety of prototypes are then presented around various proxemic relationships: from person to device and device to device interaction, and from considering factors such as f-formations and micromobility. The talk closed with some brief pointers to related work, and by walking through selected challenges within the area.

4 Overview of Talks

4.1 Investigating the Influence of Culture on Proxemic Behavior for Humanoid Robots

Elisabeth André (Universität Augsburg, DE)

License  Creative Commons BY 3.0 Unported license
© Elisabeth André

In social robotics, the behavior of humanoid robots is intended to be designed in a way that they behave in a human-like manner and serve as natural interaction partners for human users. Several aspects of human behavior such as speech, gestures, eye-gaze as well as the personal and social background of the user need therefore to be considered. In my talk, I will focus on interpersonal distance as a behavioral aspect that varies with the cultural background of the user. I will present two studies that explore whether users of

different cultures (Arabs and Germans) expect robots to behave similar to their own cultural background. The results of the first study reveal that Arabs and Germans have different expectations on the interpersonal distance between themselves and robots in a static setting. In the second study, we use the results of the first study to investigate the users' reactions on robots using the observed interpersonal distances themselves. Although the data of this dynamic setting is not conclusive, it suggests that users prefer robots that show behavior that has been observed for their own cultural background before.

4.2 Virtual Proxemics

Jakob E. Bardram (IT University of Copenhagen, DK)

License  Creative Commons BY 3.0 Unported license
© Jakob E. Bardram

In this talk I want to discuss the concept of “Virtual Proxemics”. I want to base this discussion on an on-going research project and I’m looking for input and options for brainstorming on this case together with the participants of the Dagstuhl seminar. The background for the concept of Virtual Proxemics is a research project on supporting global software development (GSD). Today, GSD is extensively used in all sorts of organizations and for all sorts of software engineering projects. Starting with the outsourcing of more trivial IT tasks like operations, support, and testing, software design, implementation, and engineering is increasingly being outsourced to countries with cheaper labor and a larger resource pool (like India, China, Pakistan, Philippines, Kenya, etc.). It is extremely well documented that GSD comes with a long list of challenges, which are related to the distributed nature of software development. In our project we operate with distance in terms of time, space, and culture. In order to mitigate these challenges and to manage large distributed software engineering projects, many organizations are using more traditional, classic waterfall-like software development methodologies – which have their own set of challenges, and often lead to project that deliver the “wrong” system later and over-budget. Agile software development methods like extreme programming (XP) and scrum have successfully been applied and have mitigated the problems of the classic software engineering problems. At the core of all agile methods is the insistence on working closely together in a collocated team of programmers, testers, product owners, and client representatives. In other words, the engine of agile development is close proximity of team members and the various tools they use. Several researcher and practitioners have asked the questions if agile methods like scrum could work in a distributed manner in a GSD project and this has been tried out in many research projects and companies. In this project, we have been working with – and studying – a company in Copenhagen that tries to apply scrum in a GSD setup with developers in India. Some of the scrum principles work, but mainly because the remote (seen from Copenhagen) team in India is represented with a local proxy, i.e. a senior lead programmer located in the Danish office. Currently we are engaging in a design process aimed at designing tools for supporting global scrum. In particular we’re interested in supporting the proxemics of a local scrum virtually over distance. A concrete design challenge is to provide the feeling of proximity across a team that is spread across (at least) two locations. Specifically we’re right now designing a global scrum board. The scrum board is the central artifact in scrum that hold all information on the progress of the project (and the so-called “sprints” in which all the work is done). The scrum board is a very public and very visible board with all sorts of information mostly written on post-it notes that are moved around. The board also work as the central focus point during the

daily scrum meetings. We want to design a scrum board that support at least the following three aspects in a virtual setup:

1. Tangible handling of post-it notes (or similar)
2. “Collocated” awareness of the progress of work
3. Ad-hoc meeting support based on the proximity of the (distributed) team in front of the board

As said, we’re in the process of designing this, and I would very much like to seek input from the participants on the seminar. I would also like to discuss the concept of “Virtual Proxemics” in greater details.

4.3 Social Interaction in Palliative Care

Susanne Boll (Universität Oldenburg, DE)

License  Creative Commons BY 3.0 Unported license
© Susanne Boll

Palliative care is taking care of individuals who are in the last year of their life. In this special phase of the life, social contacts are of important relevance for the quality of life. In these days, friends and family are often living at different places and communication is realized by selected explicit communication such as phone calls. In the same way implicit, non-verbal communication plays an important role to communicate a sense of integration into a social community. One challenge for Proxemics in HCI is to develop a sense of social proximity between geographically distant people through human-computer interaction. In our work we focus on the revival of social interaction through intuitive implicit communication and fully integrated into everyday activities. Novel multimodal human-computer interaction methods need to be designed to adapt to the individual situations of the interaction partners. In our research, we examine how through different sensory modalities such as light, sound, and by the activation of existing devices in the home people can be in implicit communication. With everyday pervasive interaction devices, which are unobtrusively integrated into the budget, we aim to raise awareness of the situation and activities locally separated but emotionally closely related individuals should be created. Simultaneously simple and intuitive ways to signal situations are recognized, in which an explicit communication channel can be initiated.

4.4 Tangible Views into Rich Information Spaces in Proximity of Large Displays

Raimund Dachsel (TU Dresden, DE)

License  Creative Commons BY 3.0 Unported license
© Raimund Dachsel

With regard to large interactive surfaces, such as tabletops and display walls, interaction research mostly focusses on two dominant ways of interaction. On the one hand, this is direct interaction on the surface of the displays, e.g., by means of multi-touch or pen input. On the other hand, it is a remote operation of a large display by users standing or sitting in some distance, e.g., by means of handheld mobile devices or mid-air gestures. We have explicitly investigated the large cubic interaction space in front of a wall display or above a tabletop and the way how rich information spaces can be mapped into this virtual volume. By means of interacting with handheld magic lenses, i.e. tangible displays tracked in space,

several users are able to explore 2D or 3D information in a very natural and seamless fashion. The complete unification of output and input space as well as the careful usage of spatial relationships between several users, several tangible displays, large contextual displays, and the virtual information allow for rich and expressive ways of navigating and exploring data spaces.

4.5 The Meaning of Space in Interaction

Joern Hurtienne (Universität Würzburg, DE)

License © Creative Commons BY 3.0 Unported license
© Joern Hurtienne

Space means a lot to us humans. In early childhood we have learned important connections between space and abstract concepts. For example, the dimension near-far is loaded with experiences in our everyday lives. We put things near to us when we need to ponder about them and put them further away when we don't (considered-is-near mapping). We group similar things close together and keep dissimilar things separate and further away (similar-is-near mapping). Friends may be physically near, but could be described as being close to us when living several thousand miles away (intimacy-is-closeness mapping). Near objects can exert their influence on us and we can exert influence on them better than on far objects (strength of effect is closeness).

We can extend these observations of so-called primary metaphors to other dimensions of space: centre-periphery, up-down, front-back, left-right, being inside or outside of containers. The questions to be discussed can be of a theoretical nature: Can we enhance Hall's ideas about proxemics with a discussion of primary metaphors? Can playing with primary metaphors in interaction design be the source of magic in using technology (e.g. as telematics breaks the everyday experience of strength-of-effect-is-closeness by letting us exert influence on distant objects). The practical goal could be to discuss specific primary metaphors and come up with lo-fi prototypes to study the proxemic effects of technology.

4.6 Hybrid-Image Visualization – or Perception-Based Proxemic Interaction

Petra Isenberg (INRIA Saclay – Île-de-France – Orsay, FR)

License © Creative Commons BY 3.0 Unported license
© Petra Isenberg

At this year's InfoVis conference we presented a first investigation into hybrid-image visualization for data analysis in large-scale viewing environments. Hybrid-image visualizations blend two different visual representations into a single static view, such that each representation can be perceived at a different viewing distance. They can be used, in particular, to enhance overview tasks from a distance and detail-in-context tasks when standing close to the display. As such, the technique allows for proximity-dependent (person to screen) interaction through locomotion and perceptual changes alone – without tracking viewers. One main question that arises is how this affects cognition (i.e. understanding and thinking about the data that is being shown) in scenarios in which people are co-located but actually see different things at the same time.

In this talk I will situate the method within the context of other techniques that show information in the same space for different viewing distances, show examples of hybrid-image visualizations, and discuss the question of cognition in more detail.

4.7 Information Visualization and Proxemics: Design Opportunities and Empirical Findings

Mikkel R. Jakobsen (University of Copenhagen, DK)

License  Creative Commons BY 3.0 Unported license
© Mikkel R. Jakobsen

People typically interact with information visualizations using a mouse. Their physical movement, orientation, and distance to visualizations are rarely used as input. We explore how to use such spatial relations among people and visualizations (i.e., proxemics) to drive interaction with visualizations, focusing here on the spatial relations between a single user and visualizations on a large display. We implement interaction techniques that zoom and pan, query and relate, and adapt visualizations based on tracking of users' position in relation to a large high-resolution display. Alternative prototypes are tested in three user studies and compared with baseline conditions that use a mouse. Our aim is to gain empirical data on the usefulness of a range of design possibilities and to generate more ideas. Among other things, the results show promise for changing zoom level or visual representation with the user's physical distance to a large display. We discuss possible benefits and potential issues to avoid when designing information visualizations that use proxemics.

4.8 Proxemics for ad-hoc communities of devices

Hans-Christian Jetter (University College London, GB)

License  Creative Commons BY 3.0 Unported license
© Hans-Christian Jetter

My goal is to use proxemic interactions between multiple devices and users to create an ad-hoc community of devices that serves users as a single usable and seamless UI. All devices of the community are aware of each other's presence and contribute their individual input/output capabilities for the common goal of providing users with a seamless, usable, and accessible interface that spans across device boundaries. Ideally, this is achieved by letting the UI's behavior emerge from simple proxemic rules that react to changes in presence, location, distance, orientation, and movement of neighboring devices and users. By using simple rules of proxemic interactions between devices, deterministic preciseness of classic top-down design and modeling is traded in against less controllable, but more adaptable, robust, and scalable bottom-up designs that automatically react to the dynamics of ad-hoc real-world usage. This will lead to self-organizing user interfaces. In this context, I also want to suggest that the more we are talking about device-to-device interactions, the less Hall's theories of proxemics help to describe the nature of interactions. David Kirsh's work on "the intelligent use of space" [1] and distributed cognition might serve as helpful frameworks here.

References

- 1 Kirsh, D. The Intelligent Use of Space. *Artificial Intelligence*, Vol. 73, Number 1–2, pp. 31–68, (1995).

4.9 Technical Challenges of Sensing People's and Devices' Proxemic Relationships

Nicolai Marquardt (University College London, UK)

License  Creative Commons BY 3.0 Unported license
© Nicolai Marquardt

This mini talk reviewed diverse sensing approaches for tracking people's and devices' proxemic relationships in ubicomp environments. The talk began with a summary of the proximity toolkit, which facilitates programming of proxemic interaction systems by providing higher-level programming building blocks. Programmers subscribe for proxemic events they are interested in, and then receive notifications about changes (e.g., a person moves closer to a particular device) through the event-driven architecture. The talk then raised the question of how we can build proxemic-aware systems without relying on high-end motion capturing systems. As one possible solution, a method of hybrid sensing is introduced, that combines (a) tracking data from structured light depth-sensing cameras, (b) radio-signal based distance sensing, and (c) the internal 6-DOF sensors. This hybrid sensing approach provides reliable tracking information; demonstrated with a series of cross-device interaction techniques. Finally, the talk raised a series of possible topics for discussions in the breakout session following the talk: (1) What are new and emerging tracking technologies, (2) what kind of tracking fidelity do we need, (3) what are adequate prototyping building blocks, and (4) what are feasible approaches for sensor fusion.

4.10 Proxemics as Play Resource

Florian Floyd Mueller (RMIT University – Melbourne, AU)

License  Creative Commons BY 3.0 Unported license
© Florian Floyd Mueller

Proxemics thinking has previously been applied to make interactions with computers more efficient. However, from computer games we know that making interactions 'harder' can result in engaging challenges. I propose that we can use proxemics thinking to contribute to our understanding of the design of challenges for digital play. In particular, I propose we can learn from related concepts in sports, where spatial relationships between players such as body contact, can make a core element of an engaging experience. By seeing proxemics as a design resource for digital play, I argue novel user experiences can be created, expanding the range of engaging interactions we experience with technology.

4.11 Situative Space Model – for human-centric ad-hoc smart environments

Thomas Pederson (IT University of Copenhagen, DK)

License  Creative Commons BY 3.0 Unported license
© Thomas Pederson

Joint work of Pederson, Thomas; Janlert, Lars-Erik; Surie, Dipak
Main reference T. Pederson, L.-E. Janlert, D Surie, "A Situative Space Model for Mobile Mixed-Reality Computing," IEEE Pervasive Computing Magazine, 10(4):73–83, 2011.
URL <http://dx.doi.org/10.1109/MPRV.2010.51>

In this talk I will introduce a body-centric modeling approach, the Situative Space Model (SSM), for mobile mixed-reality environments and relate it to the five dimensions of Ubicomp

for Proxemics proposed by Greenberg et al. (2011) [1]. The SSM is heavily influenced by proximity and divides the space close to the users into two overlapping regions: the perception space and action space, effectively defining what a human agent can perceive and act on in a given situation. Drawing from the vision of Egocentric Interaction (Pederson et al., 2010) [2] it includes real-world everyday objects (not just interactive devices) and can cope with mobility of human agents better than more device-centric approaches. The model is intended to be used both as a tool for analyzing existing mixed reality settings as well as a tool for design. ment (bibtex-files are not supported):

References

- 1 Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. 2011. Proxemic interactions: the new ubicomp?. *interactions* 18, 1 (January 2011), 42–50.
- 2 Thomas Pederson, Lars-Erik Janlert, and Dipak Surie. 2010. Towards a model for egocentric interaction with physical and virtual objects. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI'10). ACM, New York, NY, USA, 755–758. DOI: 10.1145/1868914.1869022

4.12 User-Aware Devices: How Do we Gracefully Manage Imperfect Automation?

Stacey D. Scott (University of Waterloo, CA)

License  Creative Commons BY 3.0 Unported license
© Stacey D. Scott

A key aspect of the move towards “Proxemic Interactions” is an increasing reliance on “smart” systems to track user’s body movement in order to infer user intention, and provide more responsive, and ultimately, more user-friendly systems. The underlying philosophy of this design approach is to create systems that actively collaborate with the user to provide an environment in which the technology adopts common social norms, such as increased engagement as a user approaches the system, to leverage existing knowledge of human-human interaction to improve learnability and usability with such systems. However, currently available automation (e.g., sensing technologies, algorithms for gesture interpretation, etc.) are imperfect, and failures to appropriately infer the user’s intention can increase frustration, and degrade the overall user experience. This talk with briefly overview emerging user-aware devices and pose questions for discussion about how we, as technology designers, can design our systems to gracefully handle, and allow the user to gracefully manage, such inevitable automation failures with the aim to improve the overall user experience, and overall utility and acceptability of systems that provide “Proxemics Interactions.”

4.13 Resizable Mobile Devices for Ad-Hoc Mobile Meetings

Jürgen Steimle (MPI für Informatik – Saarbrücken, DE)

License  Creative Commons BY 3.0 Unported license
© Jürgen Steimle

In this talk I propose a novel class of mobile devices to provide better support of ad-hoc mobile meetings. Advances in flexible displays will make resizable devices possible that are lightweight and have a compact form factor, while providing a quite large interactive surface when unfolded or rolled out. A jointly held large surface will allow for novel collaborative

usages in mobile settings. Taking proxemics and F-formations as a conceptual basis, I will explore several dimensions of the design space of such “handheld tabletop” devices. I will illustrate these thoughts by means of a first prototype.

4.14 Opportunities for Intelligibility in Proxemic Interactions

Jo Vermeulen (Hasselt University – Diepenbeek, BE)

License  Creative Commons BY 3.0 Unported license
© Jo Vermeulen

In this talk, we discuss opportunities for intelligibility to improve interaction with proximity-aware systems. Intelligibility could help users know how to interact with a system and know what to expect. Systems could inform users of their interactive capabilities, reassure them by highlighting mechanisms to repair mistakes and help them to anticipate the consequences of their actions. We explore possible interaction problems in proxemic interactions and discuss how different types of intelligibility could address these problems.

4.15 Siftor: subtle interaction in an art gallery context

Daniel Vogel (University of Waterloo, CA)

License  Creative Commons BY 3.0 Unported license
© Daniel Vogel

I will describe Siftor, a system that uses subtle body movement to interact with a minimal wall-sized visualization of thousands of two-dimensional art works. Using overhead cameras, the system translates the location and speed of multiple visitors into different individual and collaborative interactions. The visualization and interaction design facilitates the serendipitous discovery of art works in a conventional gallery-viewing context. Siftor was recently exhibited for seven weeks at the Owens Art Gallery in New Brunswick, Canada. The key idea is that Siftor functions as an art installation as well as a longitudinal study of interaction. For example, the simple tracking algorithm is designed to be highly flexible and permissive, making it possible to observe natural styles of body input and different strategies to master the system’s input language. Analysis of the usage logs is ongoing, but I will share initial findings relating to general usage patterns and proxemic interactions between visitors. My experience is that digital art installations in an art gallery context are well suited to conducting research in novel interaction.

Video documentation: http://www.youtube.com/watch?v=b7j_T9xUNNI

5 Working Groups

5.1 Play and Proxemics

Florian Floyd Mueller (RMIT University – Melbourne, AU)

License  Creative Commons BY 3.0 Unported license
© Florian Floyd Mueller

Joint work of Mueller, Florian Floyd; Greenberg, Saul; Dippon, Andreas; Stellmach, Sophie; Boll, Susanne

This working group discussed the topic of “play” in relation to proxemics interactions and what one field can contribute to and benefit from each other.

The session started with the group playing a (non-digital) game in the Dagstuhl environment, using locally available materials, highlighting that opportunities for play exist almost anywhere. With this experience and associated knowledge the group assembled a set of mindmaps on the topic, resulting in a following key themes:

1. How could proxemics be used for gaming?
 2. Proxemics appears to have potential to engage people into play: we should envision strategies informed by proxemics
 3. Proxemics awareness could trigger curiosity
 4. One strategy could be to visualize proxemics, this appears to be affording playfulness
 5. Connecting space by play could be another strategy
 6. Location-based games seem to focus on absolute distances, proxemics on relative distances
 7. Proxemics play is related to new dating apps that take location into account: they often ask “is there a compatible match nearby?”
 8. Playing with pictures can also benefit from proxemics: “what pictures have been taking nearby my location?”
1. Proxemics games to help teach social issues
 2. Proxemics in games could take on a supportive role to create the possibility to deal with particular problems or issues, such as the fear to interact with strangers, deal with anger when losing, promote more rapid intimacy, teach social behavior, make social behavior more explicit, etc. One example system could be a musical chair or cocktail glasses that function as social mixer.
1. Digitally exaggerating proxemics could be fun we believe
 2. Twister has been described as a game where the bodies are the play pieces, this could be inspirational for proxemics play
 3. An underexplored area seems to be proxemics ≤ 0 cm, often called contact sport in sports contexts
1. The “Magic circle of play” (Zimmerman and Salen, 2003) could be a frame for proxemics
 2. Interaction designers could play with the social norms and expectations of the magic circle of play by using proxemics as frame

The group decided to work collaboratively on a publication on the topic, with the DIS (Designing Interactive Systems) conference as a possible target conference.

5.2 Intelligibility for Proxemic Interactions

Jo Vermeulen (Hasselt University – Diepenbeek, BE)

License © Creative Commons BY 3.0 Unported license
© Jo Vermeulen

Joint work of Boll, Susanne; Boring, Sebastian; Dachselt, Raimund; Dostal, Jakub; Isenberg, Petra; Marquardt, Nicolai; Matulic, Fabrice; Mueller, Florian Floyd; Nicosia, Max; Reiterer, Harald; Scott, Stacey; Vermeulen, Jo; Vogel, Daniel

Introduction

In this breakout group, we wanted to delve deeper into interaction issues of systems that rely on proxemics. While many of the systems described in the literature work well for a specific setting, the implicit nature of proxemic interactions could also cause problems for users. During the introductory talks at the start of the seminar, several of these issues were raised, including: discoverability, providing control, graceful failure and correcting mistakes, correctly detecting users' intentions, how to know what is being tracked by the system, or how to opt-out (and avoid unintentionally interacting with the system). As mentioned by Greenberg et al. [5], many of the suggested interaction techniques assume the existence of a set of rules of behavior that dictate what the different entities should do based on implicit acts. There will be many situations in which applying these rules will be the wrong thing to do [6]. Additionally, Ballendat et al. [1] argue that one of the largest unsolved issues in proxemic interactions is how one can configure the rules of behavior, and how users can repair mistakes when the system gets it wrong.

A number of these problems have been reported earlier and are well known in the area of context-aware computing [4]. One of the major challenges with context-aware systems is making these systems intelligible [2] by informing users about the system's understanding of the world. A proxemic relationship between devices and people is essentially nothing more than a specific type of context information that can be taken into account. The breakout group was interested to see whether existing approaches to address these challenges in context-aware systems (e.g., mediation [3]) could also be applied to systems that take into account proxemics, and possible ways to tackle issues specific to proxemic interactions. For example, Ju et al. [7] propose three interaction techniques that could be used to show users what the system is doing (system demonstration), how the users input is interpreted (user reflection) or to correct the system when it makes a mistake (override).

Purpose of the Breakout Group

We set out to discuss discovery, mediation, intelligibility and visualizations in proxemic interactions. Given time limitations, we decided to focus specifically on the problems of opting in and opting out. We also looked into how the user's level of engagement with the system could help in addressing these challenges. As the group was too large for a single breakout session, we split up in two subgroups: (1) opt-in and (2) opt-out / methods of engagement.

Participants of the Breakout Groups

Contributors to both subgroups are listed below.

Opt-in

- Susanne Boll, Sebastian Boring, Raimund Dachselt, Harald Reiterer, Stacey Scott, Jo Vermeulen, and Daniel Vogel.

- The notes for this subgroup are listed in Appendix 1. Daniel illustrated several ideas using sketches. Jo was the group scribe.

Opt-out / Methods of Engagement

- Jakub Dostal, Petra Isenberg, Nicolai Marquardt, Fabrice Matulic, Florian Floyd Mueller, and Max Nicosia.
- The notes for this subgroup are listed in Appendix 2. Nicolai and Fabrice created sketches to capture the discussion.

Outcome

Both subgroups set out to identify the specific problems that occur when opting in or opting out, propose potential solutions to these problems, and discuss how those solutions would be applied in specific scenarios. Both groups used sketches to capture and illustrate their ideas. The brainstorming resulted in many interesting – although somewhat rough – ideas. In the following, we provide a brief overview of recurring themes and concepts.

Social Protocol. One of the proposals, discussed at length in the opt-in subgroup, was the idea to make systems adhere and respond to a social protocol, just like in human-to-human communication. There are several subtle clues in our day-to-day communication that tell others whether we are approachable or not and allow us to opt-in or opt-out (e.g., a brief nod and smile to someone you recognize at the other end of the room at a reception, looking away to avoid starting a conversation).

Multiple Levels of Opt-in. The opt-in subgroup asked the question whether a single level of opt-in might be insufficient. It could be useful to support several levels of opt-in and opt-out (e.g., depending on distance, eye-contact, orientation). Users familiar with the system could then be automatically opted into a deeper level in the hierarchy.

Opt-out Gestures. The opt-out subgroup brainstormed about different possible gestures to opt-out. Several ideas were proposed, such as a 'stop' gesture, covering your face with your hands, or turning away from the system. Similarly, the opt-in subgroup asked the question what would be the proxemics equivalent of sticking a post-it note over a laptop's webcam to avoid being tracked, and which would be more obtrusive, being filmed or having ones proximity tracked? The use of special clothing to indicate willingness to opt-in (e.g., a shirt in a specific color, wearing a special type of hat or cap) was also discussed, as well as the idea of sensible opt-outs, where a system could, for example, avoid implicitly opting in small children.

Ownership of the Space and Interaction Zones. Both groups discussed issues related to how proxemics-aware systems use the space in which they are deployed. A question raised in the opt-in subgroup was who owns the space surrounding the system, as the deployment of the system could influence how people use that space. Users might, for example, need to walk around a public display to avoid opting in, making the space around it more crowded, which could be problematic at rush hour. The opt-out subgroup proposed using a special entrance to opt-in to the system, instead of using interaction zones defined by distance.

Methods of Engagement. The second subgroup considered how the user's level of engagement with the system could be used. Systems could rely on more implicit or explicit means of interaction. For example, they could react to users' presence when they are just walking by, or might require the user to approach the system as a more explicit signal of intent. Similarly, the opt-in subgroup discussed how interaction with physical props near a

public display could be used as an explicit way of opting in. To indicate what is tracked and how the system responds to user input (e.g., distance, eye gaze, orientation), this subgroup proposed the use of specific icons or signs.

User Control. Participants also discussed how users can be allowed to exert control over the system, and in what situations a lack of user control could be problematic. Implicitly tracking proxemic dimensions such as distance, orientation or movement and reacting to changes in these dimensions can have annoying side effects. For example, a public display that uses distance to control the zoom level does not allow users to step closer and get a detailed view of a specific part of its contents, as the display will adjust the zoom level in response to their approach. An idea proposed by the opt-in subgroup was the use of a symmetric opt-in and opt-out process, in which opting out could be done by performing the inverse of the opt-in action (e.g., if users would opt-in by approaching the system, they could opt-out by moving back).

References

- 1 Till Ballendat, Nicolai Marquardt, and Saul Greenberg. Proxemic interaction: designing for a proximity and orientation-aware environment. *ACM International Conference on Interactive Tabletops and Surfaces*, ACM (2010), 121–130.
- 2 Bellotti, V. and Edwards, K. Intelligibility and Accountability: Human Considerations in Context-aware Systems. *Hum.-Comput. Interact.* 16, 2 (2001), 193–212.
- 3 Dey, A.K. and Mankoff, J. Designing Mediation for Context-aware Applications. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (2005), 53–80.
- 4 Dey, A.K. Understanding and Using Context. *Personal Ubiquitous Comput.* 5, 1 (2001), 4–7.
- 5 Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. Proxemic interactions: the new ubicomp? *interactions* 18, 1 (2011), 42–50.
- 6 Greenberg, S. Opportunities for Proxemic Interactions in Ubicomp (Keynote). In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2011*. Springer Berlin Heidelberg, 2011, 3–10.
- 7 Ju, W., Lee, B.A., and Klemmer, S.R. Range: exploring implicit interaction through electronic whiteboard design. *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, ACM (2008), 17–26.

Appendix 1: Notes of Opt-in Subgroup

- Levels of opt-in
 - First level of going into space
 - Delineated space
 - 1 pixel bar Dan
 - * Pure opt-in? People might never do it
 - * Continuous feedback (something moves when I move)
 - Know about system, context matters
 - * Maybe I just want to use the whiteboard in an analog way
- Automatic opt-in for feedback
- Just want to look at the map (work of Mikkel)
- Scenario?
- Tile → floor changes
- Social protocol

- Compare with human-human behaviour
 - * Look at someone, opt-in, to start a conversation
 - Level of invitation (Dan)
 - * People that look away
 - Displays always want to invite people in (assumption)
 - * Allowing user to quickly figure out if display relevant to them (Stacey)
 - * Even advertisers don't want everyone to look at it → targeted
 - * Encourage approach
 - * Who are we?
 - Act in social way
 - Natural ways of opting-out, by just doing the opposite
 - Different levels, teleportation into deeper level if you're an expert
- Does it have to be implicit or automatic?
 - What can you model, what don't you need to model?
 - * Use of physical props, car → when you open door → show display
 - Do we need to use zones? We can't just use zones alone, we need more information
 - * But we can't use lots and lots of sensors, some things cannot be modelled
- Physical way to opt-out
 - Sticker on webcam to opt-out
 - Flash camera against pictures
 - What for proximity?
 - How acceptable are different sensors
 - * Proximity vs. camera?
- Stacey: overlap with territoriality (who owns the space, you occupy)
 - Primary, secondary, public (degrees of defending territory)
 - In public: explicit opt-in
 - Multiple users: one person opt-in, other one is standing beside them, still interact together
 - Social correctness: moving furniture in people's places, chair, table, etc. different
- Shopping scenario
 - Shelf where you can put products on
 - Display that shows info
 - Shopping cart could be mediator
 - Push information to private display
 - Compare in shopping cart
 - At what point do you identify yourself
 - * Opt-in with your position compared to products (passing by with shopping cart)
 - Compared to shopping cart
 - Implicitly opting in
 - Push private info to display
 - Show info on phone
 - Bring products near you (shopping cart)
 - Compare them on shopping cart display
 - * Explicitly opting in
 - Social protocol

- * Object flashing, like recognizing you, waving hand
- * Close the flap when you don't want to be bothered
 - Explicit opt out
- * Depending on available time / shopping type
 - Everyday shopping vs. explorative shopping
 - Finding products

Appendix 2: Notes of Opt-out / Methods of Engagement Subgroup

- Novice versus an expert
- Explicit vs. implicit Opting in vs. action Transition between the two
- The environment may offer different levels of engagement
 - Walk in a space and you are immediately being sensed vs.
 - Come near a device and do something more explicit to do something with it
 - Different phases probably have smooth transitions from passing by to direct interaction and back to leaving
- Possible ways to indicate actions
- How to inform the user of what is being tracked
- Opt-out / opt-in
 - What is the default?
 - What is the cost of opting in or opting out
 - Being overwhelmed
 - Continuously increasing actions to opt-out
- Links
 - Tracking customers in stores using WiFi:
<http://www.washingtonpost.com/blogs/the-switch/wp/2013/10/19/how-stores-use-your-phones-wifi-to-track-your-shopping-habits/>
 - Tesco face recognition for targeted ads:
<http://news.sky.com/story/1163551/tesco-face-scanners-to-target-till-adverts>
 - Infrared masks to blind cameras:
<http://mods-n-hacks.wonderhowto.com/how-to/make-infrared-mask-hide-your-face-from-cameras-201280/>
 - Wi-Fi Beacons – Prove request details (Device tracking):
<http://www.wi-fiplanet.com/tutorials/print.php/1447501/>

5.3 Dark Patterns in Proxemic Interactions

Saul Greenberg (University of Calgary, CA)

License  Creative Commons BY 3.0 Unported license
© Saul Greenberg

Joint work of Boring, Sebastian; Dostal, Jakub; Greenberg, Saul; Isenberg, Petra; Matulic, Fabrice; Pederson, Thomas; Scott, Stacey; Vermeulen, Jo

Introduction

Authors of human-computer interaction papers concerning innovative design ideas tend to forward their central idea in a positive – often highly idyllic – light. True critical perspectives are rarely offered. When they are, they tend towards a few cautionary lines in the discussion,

or are relegated to future work where its actual use would be examined. The problem is that many of our new innovations involve designing for ubiquitous computing situations that are extremely sensitive to intentional or unintentional abuse (e.g., privacy, distraction and intrusion concerns). Rather than wait until some future field study of our technology (where it may be too late to address emerging concerns), we should consider the 'dark side' of our technologies at the outset.

The particular innovation we are concerned with in this Dagstuhl Workshop is proxemic interactions, which was inspired by Hall's Proxemic theory. The theory explains people's understanding and use of interpersonal distances to mediate their social interactions with others. In proxemic interactions, the intent is to design systems that will let people exploit a similar understanding of their proxemic relations with their nearby digital devices to facilitate more seamless and natural interactions. This is especially important as we become immersed in ubiquitous computing ecologies, i.e., where we carry and are surrounded by myriads of devices, all potentially capable of interacting with one another. Examples include: mobile devices that understand their spatial relations to mediate information exchange between nearby devices; large displays that sense people's position relative to them, where they dynamically adjust what is shown and how people can interact with them; public art installations that respond to the movement and proximity of people within its sphere to affect what is shown; application areas such as home media players that monitor the distance and orientation of its viewers to dictate what is shown, and information visualizations that tune their visuals to people's position relative to them. The literature also includes more general essays about the role of proxemics, such as how it can address well-known challenges in Ubiquitous Computing design.

Yet it is clear, at least intuitively, that there is a dark side to proxemics interactions. For example, the systems above rely on sensing people and their devices within the surrounding environment. We already know that some of the sensed dimensions that would be valuable to proxemic system design include: distance, orientation, and movement of entities relative to one another, the identity of these entities, and contextual information about the location. While the purposes of researchers within this area are honorable, such sensing immediately raises concerns about privacy by experts and non-experts alike. Moreover, dystopian visions of the future hint at abuses of such technologies – a well-known example is the movie *Minority Report* that illustrates how a character is bombarded by targeted advertisements as he moves through a public hallway, and how his location is revealed to searchers.

The Purpose of the Breakout Group

In this breakout group, we revisited the idea of proxemic interactions, where our goal was to discuss a critical perspective – the dark side – of this technology. Our method was to articulate potential dark patterns indicating how we think this technology can be – and likely will be – abused, and anti-patterns where resulting behavior occurs as an unintended negative side effect. Participants articulated not only possible deceptions and misuses of proxemics interactions (dark patterns), but problems that may appear even when the designer has reasonable intentions (anti-patterns).

Unlike true patterns that are based on analyzing a broad variety of existing solutions, we brainstormed patterns based on several sources. As part of our investigation, we revisited Brignull's dark patterns web site (darkpatterns.org) to see if and how the dark patterns recognized in web browsing systems could be applied to proxemic interactions (possibly as variations). We also considered emerging uses of proxemics in commercial and experimental

products as examples, as well as 'thought experiments' of how such systems could be designed with dark patterns in mind. We considered dark portrayals of such technologies foreshadowed in the popular literature and cinema, and our own reflections of where misuses could occur. That is, our patterns are a mix of those that describe existing abuses and that predict possible future ones. We did not differentiate whether a particular pattern is dark vs. anti: our pattern examples suggested that the difference between the two often arise from the designer's intent rather than a feature of a particular design. In our view, the same pattern – depending on the designer's intent – can be viewed as either a dark pattern or an anti-pattern. We believe this approach to be appropriate for forecasting – and ideally mitigating – the dark side of our future technologies before actual deceptive patterns become widespread in practice.

A set of initial patterns and the notes accompanying them are attached as Appendix 1.

Participants of the Breakout Groups

Contributors to the breakout group were:

- Kakub Dostal, Fabrice Matulic, Jo Vermeulen, Petra Isenberg, Saul Greenberg, Sebastian Boring, Stacey Scott, and Thomas Pederson
- Petra Isenberg was the group scribe, where the outcome of her work is listed in Appendix 1.
- Aaron Quigley, while not part of the breakout group, deserves special mention as he primed the group with the dark pattern web site when we initially discussed the dark side of proxemic interactions.

Outcome

During the workshop, there was sufficient interest by attendees to develop this idea as a paper.

In early December, Greenberg developed the framework of a paper and wrote a few sections, along with Boring (who happened to be visiting). He then asked who in the group would be willing to participate in the development of the paper, with the proviso that they would be active authors. This was, in part, because Greenberg and Boring were targeting the ACM DIS 2014 conference, where the paper submission deadline was in mid-January. Of the original participants, Jakub Dostal, Jo Vermeulen, Saul Greenberg and Sebastian Boring agreed to be active authors, and others said they were happy to comment on it. As of time of this writing, a complete draft of a paper has been prepared and is available as a technical report [1].

A paper based on this report has been submitted to a conference. The paper itself is a substantial reworking of the original patterns as brainstormed in Appendix 1. That is, the breakout group was excellent in terms of motivating the theme of Dark Patterns and in brainstorming initial discussion. However, as with the results of most brainstorming exercises, it demanded considerable effort to transform these initial thoughts into a publishable form.

References

- 1 Greenberg, S., Boring, S., Vermeulen, J. and Dostal, J. (2014) Dark Patterns in Proxemic Interactions: A Critical Perspective. Research report 2014-1055-05, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada, January. <http://grouplab.cpsc.ucalgary.ca/Publications/2014-DarkPatterns.Report2014-1055-05>

Appendix 1: Initial Patterns brainstormed by the Group

The group brainstormed the following initial set of patterns and notes for each pattern. Design patterns were initially taken from darkpatterns.org and more or less adapted to Proxemics, as well as some new patterns proposed.

Bait and switch

- by approaching something you are automatically opting in (implicit action)
- benevolent solutions: needs to be possible to opt out.
 - 1) Gesture for opting out?
 - 2) if you notice that you are tracked, maybe your reaction can be interpreted as “i don’t want to be tracked”
- trust is crucial
- cameras make you think you are tracked even if you are not – changes your behavior
- “for implicit things, only safe actions”
- “we no longer own the space”

Forced behaviour

- people are forced to a certain (embarrassing) behaviour in order to use service
- tricks you into go closer (interesting!) then you are forced to see an ad/pay

Disguised ads (disguised tracking) implicit consent

- ads in a public space
- make you get close enough for a picture, then target your face for future ads

The captive audience

- small display on top of mens urinals (you cannot go somewhere else to avoid watching)
- “black mirror” british tv show, you need to pay to avoid the ads on the display walls
- kinect enters your living room, disguised as an entertainment system, silently tracks everything

Faraway bill

- the proxemics system forces you to go to a location in order to get a service

Forced continuity

- forced to remain and watch an ad before leaving the space or all your data will be deleted from the cloud
- when two devices are brought together they share data with their owners consent. next time they will continue to share data, even without the consent

Forced disclosure

- everything is taken from your mobile device as you approach the system

Friend spam

- a system might automatically connect you to people you happen to be close to
- viral: a “friend virus” that spreads among people you are physically meeting

Hidden costs

- in the last stage of the checkout process unexpected charges occur
- proxemics case: you use the service and then you are asked to pay (somehow), e.g. with time
- might work if the fee is not so big
- like fitbit: when you are synching, you are suddenly asked to pay

Unintended relationships

- just because you pass by someone, you are automatically friended with that person

Misdirection

- animations flashing to attract your attention so that the camera can get a good picture of you
- hide information by placing the legal text in a place that is not possible to read

Privacy zuckering

- making it hard to get full privacy

Roach motel

- move up to a public display. when you leave, you need to pay in order for your private data to not stick to the screen.

Trick question

- if you layered info as you move towards the device, you get a quick question that you can't correctly interpret because you are in motion/you are not oriented correctly

Attention grabbing

- proxemics provides better timing

Midas touch problem exploited**Physical aspect**

- lure them into positions
- lure them into getting their finger print

Intentional vs. unintentional dark patterns**5.4 Ad-Hoc Proxemics – Inclusion of everyday entities in proxemics systems**

Thomas Pederson (IT University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license
© Thomas Pederson

Joint work of Butz, Andreas; Dippon, Andreas; Hurtienne, Jörn; Jetter, Hans-Christian; Sorensen, Henrik; Stellmach, Sophie; Pederson, Thomas; Rädle, Roman

This working group deliberately turned the focus away from what had been a recurring topic at this Dagstuhl seminar: pre-designed systems that make use of proximity as a means for interaction (e.g. proximity-aware public displays), and instead discussed the potential role that proximity plays when interacting with physical entities (objects) in everyday life and how designers of proximity-based systems need to take that into account.

The interplay between physical structures and the proximity-based system

Since real world objects and structures seem to influence human agents' interpretation of what can be done (and not done) in a given environment (e.g. few entities and structures in a car garage tell us that we could/should bake a cake there), interactive systems that make use of proximity (both object↔object and human agent↔object)

1. should avoid introducing proxemic behaviour that the physical environment as such does not indicate or afford,
2. should as much as possible leverage on proxemic behaviour that the physical environment indeed is signalling to the human agent as possible or even encouraged.

While the above two reciprocal requirements can be fulfilled when a proxemic-based system is set up in a controlled static environment such as a dedicated room, it becomes a challenge if proxemic behaviour is used in mobile settings where the physical surroundings in which the system operates is different from one time to the next.

The working group also came to the conclusion that even for the design of static proximity-based systems (e.g. a proximity-aware public display), it could be beneficial as system designer to take into account, and influence, the physical structure surrounding the interactive system such as to indicate to the users of the system what can be done, and how.

With respect to individual physical objects/entities, their physical properties (shape, colour, weight, rigidity) might be possible to design in such a way as to indicate if, and how, the specific object reacts on/can be used for proximity-based interaction.

The semantics of inter-object proximity

It is well documented in literature (e.g. Kirsh, 1995) that the organization of physical entities in space is associated with the meaning which the “space organizing human agent” projects onto the objects. The most evident fact is that objects that are related end up close to each other. It was concluded that any interactive system that wants to model the intentions of human agents, could benefit from taking inter-object proximity into account.

Semantics can also be built into the physical environment in such a way that human agents that operate in the environment are consciously or unconsciously led to “do the right thing”. Example: Silverware at the dining table might be placed in such a way that the spoon is only reachable when the plate is gone.

Does point of reference matter?

Does it matter whether the proximity-aware system uses the human body as center of reference or the room? It was concluded that for certain system tasks, it might matter. The decision depends on which of the approaches that provides the best view of the situation for the system. Combined viewpoints are also possible (e.g. that devices communicate with each other to better identify the situation).

Important property: Everyday objects are inexpensive and ubiquitous

An inclusion of everyday objects (such as paper documents, pens, cutlery) in interactive systems would open up for new kinds of interaction over both time and space due to the fact that their situational availability is immensely higher than typical digital devices. Part of Mark Weiser’s vision for Ubiquitous Computing relied on spreading out the inexpensive devices everywhere. If the everyday objects, to some degree, can take on the role of such devices, they do not need to be spread out because they already are!

The idea of using everyday objects as controllers for virtual/digital processes is not completely new, see for instance Henderson & Feiner (2008); Corsten et al. (2013); MaKey Makey (<http://www.makeymakey.com>). The working group identified three important roles that everyday objects could take as part of interactive systems:

- controllers
- modifiers
- mediators

The Danger / Challenges

Integrating everyday objects into interactive systems potentially makes the up until now very predictable real world suddenly much less predictable. The working group acknowledged that special care was needed in particular in environments where spontaneous encounters between human agents and objects occur (public places) whereas more private environments could be less problematic due to the fact that everyone operating in the environment will know what virtual processes are tied to what everyday object; what spatial configuration of objects will initiate what virtual process, etc. etc.

Other topics, conclusions drawn, and ideas

- Absolute proximity vs. relative proximity (object↔object and human agent↔object).
- Fine grained object manipulation doesn't necessarily demand fine grained tracking.
- The idea of virtual mobility: virtual "content" moves towards you instead of the other way round. E.g. information ends up on your personal device instead of a wall-sized display.
- The Reality-Based Interaction Framework (Jacob et al., 2008) is highly relevant to the discussion on relying on everyday proxemics for designing better interactive systems.
- For some tasks, in particular in dedicated places designed for "expert users", a high learning threshold for interacting with a proximity-based interactive system might be fully OK.

Outcome

The participants of the working group are considering to set up a workshop at an upcoming conference (for instance MobileHCI 2014) to dig deeper into some of the topics mentioned above.

5.5 Challenges of Sensing People's and Devices' Proxemic Relationships

Nicolai Marquardt (University College London, UK)

License © Creative Commons BY 3.0 Unported license
© Nicolai Marquardt

Joint work of Dachsel, Raimund; Jetter, Hans-Christian; Rädle, Roman; Sorensen, Henrik, Dostal, Jakob; Nicosia, Max; Pederson, Thomas, Dippon, Andreas, Bardram, Jakob E.

The theme of this breakout session was the discussion of existing challenges and future technical approaches for sensing people's and devices' proxemic relationships. We began by collecting common tracking technology approaches: vision-based (e.g., structure light cameras, motion capturing systems, thermal cameras), radio-based (e.g., Bluetooth, Wifi, RFID), sensor-based (e.g., infrared, ultrasonic, microphone, magnetometer). We then categorized these approaches along the low-fidelity to high-fidelity spectrum. Next step was to brainstorm characteristics and properties that are important to consider when choosing between different tracking alternatives: precision, power consumption, uncertainty, outdoor vs. indoor use, user preference, weight, cost, scalability, complexity of processing, and others. Finally, as the major part of this breakout session, we discussed strategies for combining different sensing technologies with sensor-fusion approaches. In here, alternative strategies are possible:

multiple technologies can complement each other (e.g., work at different scales) or reinforce the results of another (e.g., all tracking same area but fusion increases resolution). Other aspects important for sensor fusion approaches are: the weighting of sensors, approaches for graceful failure, hierarchical sensing approaches, sensor roaming, and translation of sensor data. As possible future outcomes of the breakout discussion we are considering the setup of a website facilitating the comparison and selection of proxemic tracking technologies (e.g., making suggestions for technology based on set of requirements).

Participants

- Elisabeth André
Universität Augsburg, DE
- Jakob E. Bardram
IT Univ. of Copenhagen, DK
- Susanne Boll
Universität Oldenburg, DE
- Sebastian Boring
University of Copenhagen, DK
- Andreas Butz
LMU München, DE
- Raimund Dachsel
TU Dresden, DE
- Andreas Dippon
TU München, DE
- Jakub Dostal
University of St. Andrews, GB
- Saul Greenberg
University of Calgary, CA
- Kasper Hornbæk
University of Copenhagen, DK
- Jörn Hurtienne
Universität Würzburg, DE
- Petra Isenberg
INRIA Saclay – Île-de-France –
Orsay, FR
- Mikkel R. Jakobsen
University of Copenhagen, DK
- Hans-Christian Jetter
University College London, GB
- Nicolai Marquardt
University College London, UK
- Fabrice Matulic
ETH Zürich, CH
- Florian Floyd Mueller
RMIT Univ. – Melbourne, AU
- Max Nicosia
University of St. Andrews, GB
- Kenton O’Hara
Microsoft Research UK –
Cambridge, GB
- Thomas Pederson
IT Univ. of Copenhagen, DK
- Aaron Quigley
University of St. Andrews, GB
- Roman Rädle
Universität Konstanz, DE
- Harald Reiterer
Universität Konstanz, DE
- Stacey D. Scott
University of Waterloo, CA
- Henrik Sorensen
Aalborg University, DK
- Jürgen Steimle
MPI für Informatik –
Saarbrücken, DE
- Sophie Stellmach
TU Dresden, DE
- Jo Vermeulen
Hasselt Univ. – Diepenbeek, BE
- Daniel Vogel
University of Waterloo, CA



Electronic Markets and Auctions

Edited by

Yishay Mansour¹, Benny Moldovanu², Noam Nisan³, and
Berthold Vöcking⁴

1 Tel Aviv University, IL, mansour@cs.tau.ac.il

2 Universität Bonn, DE

3 The Hebrew University of Jerusalem, IL, noam@cs.huji.ac.il

4 RWTH Aachen, DE

Abstract

The main goal of this seminar was to study topics related to electronic markets and auctions both from the computational perspective and from a game-theoretic and economic one. From the computer science perspective, with the advent of the Internet, there has been a significant amount of work in Algorithmic Game Theory focusing on computational aspects of electronic markets and on algorithmic aspects of mechanism design. Economics have been traditionally interested in markets in general and designing efficient markets mechanisms (such as auctions) in particular. The recent emergence of electronic markets and auctions has only reemphasized the importance of this topic.

Seminar 10.–15. November, 2013 – www.dagstuhl.de/13461

1998 ACM Subject Classification J.4 Social and Behavioral Sciences, F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Algorithmic game theory, mechanism design, economics, electronic markets

Digital Object Identifier 10.4230/DagRep.3.11.58

Edited in cooperation with Andreas Tönnis

1 Executive Summary

Yishay Mansour

Noam Nisan

License  Creative Commons BY 3.0 Unported license
© Yishay Mansour and Noam Nisan

The relatively young field of Algorithmic Game Theory sets a goal of providing a computational understanding of game theory models. The research in the field has many focal points, including exploring the quality of equilibria, computation of equilibria, algorithmic mechanism design, as well as analyzing computer science related games and gaining an economics perspective for many important optimization problems.

While it is still too early for the evaluation the long term contribution of Algorithmic Game Theory to the field of Economics, in general, and to Game Theory in particular, we would like to highlight some successful contributions. The efficient computational aspects are a clear contribution, and this is also coupled with the understanding that sub-optimal solutions can have various degrees of sub-optimality. By using approximation algorithms approaches traditional in Theoretical Computer Science, the sub-optimality can be quantify in a very rigorous and clear way. The study of concrete convergence rates, rather than



Except where otherwise noted, content of this report is licensed
under a Creative Commons BY 3.0 Unported license

Electronic Markets and Auctions, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 58–78

Editors: Yishay Mansour, Benny Moldovanu, Noam Nisan, and Berthold Vöcking



DAGSTUHL
REPORTS Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

convergence in the limit, has proved to be highly successful here, as well. Finally, the extensive study of discrete models, especially combinatorial auctions, has been an area where computer science has made significant contributions.

The economic field of Mechanism Design asks how to design mechanisms that will implement some desired social choice function under rational behavior of the participants. This field is at the forefront of economics research, and its goal is to gain a better understanding of designing mechanisms that considers the incentives of participant. This is in general viewed as part of market design, and micro-economics

One of the central areas of Algorithmic Game Theory is *Algorithmic Mechanism Design*. This field is relevant to designing distributed computer systems, suggested that mechanism design should also consider the algorithmic issues involved beyond the strategic ones commonly studied in economics. The seminar concentrated on Algorithmic Game Theory, with an emphasis on the sub-field of Algorithmic Mechanism Design.

The central application of Mechanism Design is the implementation of auctions and markets, and similarly the central application of algorithmic mechanism design is the implementation of complex computerized auctions and markets. As markets and auctions are increasingly implemented over computer networks, and as they are getting more sophisticated, much theoretical research has gone into the design of complex auctions and markets. Issues that need to be treated include computational ones, strategic ones, and communication ones. A central application is, so called, combinatorial auctions, which aim to concurrently sell many related items.

This seminar had researchers discussing basic research questions that lie behind the growing challenges in electronic markets and auctions. The seminar took a broad view of these challenges, focusing on foundational issues, taking a wide perspective, from the high-level issues of Algorithmic Game Theory through the Algorithmic Mechanism Design aspects, to basic challenges of electronic markets and auction.

2 Table of Contents

Executive Summary

<i>Yishai Mansour and Noam Nisan</i>	58
------------------------------------------------	----

Overview of Talks

Single parameter mechanism for unrelated machine scheduling <i>Yossi Azar</i>	62
On the Efficiency of the Walrasian Mechanism <i>Moshe Babaioff</i>	62
The Tradeoff between Efficiency and Strategyproofness in Randomized Social Choice <i>Felix Brandt</i>	63
Algorithms for Strategic Agents I: Revenue Maximization <i>Yang Cai</i>	63
Economic Efficiency Requires Interaction <i>Shahar Dobzinski</i>	64
Towards More Practical Linear Programming-based Techniques for Algorithmic Mechanism Design <i>Khaled Elbassioni</i>	64
A unified approach to restricted complements <i>Michal Feldman</i>	64
Online price of anarchy for parking <i>Amos Fiat</i>	65
Optimal Impartial Selection <i>Felix Fischer</i>	65
Manipulation of Stable Matchings using Minimal Blacklists <i>Yannai A. Gonczarowski</i>	65
Optimal Competitive Auctions <i>Nick Gravin</i>	66
Quantitative Comparative Statics for a Multimarket Paradox <i>Tobias Harks</i>	67
Redesigning the Israeli Psychology Market <i>Avinatan Hassidim</i>	67
Designing Profit Shares in Coalition Formation Games <i>Martin Hoefer</i>	67
Privacy-Preserving Auctions <i>Zhiyi Huang</i>	68
Duality and optimality of auctions for the uniform distribution <i>Elias Koutsoupias</i>	68
Characterization of SMON mechanisms with additive valuations over the real domain <i>Annamaria Kovacs</i>	69

Mechanisms for Multi-Unit Combinatorial Auctions with a Few Distinct Goods <i>Piotr Krysta</i>	69
Prior-free Auctions with Ordered Bidders <i>Stefano Leonardi</i>	70
Implementing the “Wisdom of the Crowd” <i>Yishay Mansour</i>	70
Deferred Acceptance Auctions <i>Paul Milgrom</i>	71
Plasticity, Monotonicity and Implementability <i>Rudolf Mueller</i>	71
Revenue Maximization with Sampling <i>Noam Nisan</i>	71
Dynamic Models of Reputation and Competition in Job-Market Matching <i>Sigal Oren</i>	72
Matchings, Vertex Cover und Network Bargaining Games <i>Britta Peis</i>	72
Learning Equilibria of Games via Payoff Queries <i>Rahul Savani</i>	73
Non Adaptive Methods for Adaptive Seeding <i>Yaron Singer</i>	73
Cost-Recovering Bayesian Algorithmic Mechanism Design <i>Balasubramanian Sivan</i>	73
Composable and Efficient Mechanisms <i>Éva Tardos</i>	74
An Optimal Online Algorithm for Weighted Bipartite Matching and Extensions to Packing Linear Programs <i>Andreas Toennis</i>	74
Local computation mechanism design <i>Shai Vardi</i>	75
Algorithms for Strategic Agents II <i>S. Matthew Weinberg</i>	75
Utility-Target Auctions <i>Christopher A. Wilkens</i>	76
A Unified Approach to Truthful Scheduling on Related Machines <i>Rob van Stee</i>	76
Participants	78

3 Overview of Talks

3.1 Single parameter mechanism for unrelated machine scheduling

Yossi Azar (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Yossi Azar

Joint work of Azar, Yossi; Maor, Idan

We show a randomized truthful mechanism for the restricted-related scheduling model with 3-approximation on the makespan.

3.2 On the Efficiency of the Walrasian Mechanism

Moshe Babaioff (Microsoft Research – Mountain View, US)

License © Creative Commons BY 3.0 Unported license
© Moshe Babaioff

Joint work of Babaioff, Moshe; Lucier, Brendan; Nisan, Noam; Paes Leme, Renato

Main reference M. Babaioff, B. Lucier, N. Nisan, R. Paes Leme, “On the Efficiency of the Walrasian Mechanism,” arXiv:1311.0924v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1311.0924v1>

Central results in economics guarantee the existence of efficient equilibria for various classes of markets. An underlying assumption in early work is that agents are price-takers, i.e., agents honestly report their true demand in response to prices. A line of research in economics, initiated by Hurwicz (1972), is devoted to understanding how such markets perform when agents are strategic about their demands. This is captured by the *Walrasian Mechanism* that proceeds by collecting reported demands, finding clearing prices in the *reported* market via an ascending price tâtonnement procedure, and returns the resulting allocation. Similar mechanisms are used, for example, in the daily opening of the New York Stock Exchange and the call market for copper and gold in London.

In practice, it is commonly observed that agents in such markets reduce their demand leading to behaviors resembling bargaining and to inefficient outcomes. We ask how inefficient the equilibria can be. Our main result is that the welfare of every pure Nash equilibrium of the Walrasian mechanism is at least one quarter of the optimal welfare, when players have gross substitute valuations and do not overbid. Previous analysis of the Walrasian mechanism have resorted to large market assumptions to show convergence to efficiency in the limit. Our result shows that approximate efficiency is guaranteed regardless of the size of the market.

We extend our results in several directions. First, our results extend to Bayes-Nash equilibria and outcomes of no regret learning via the smooth mechanism framework. We also extend our bounds to any mechanism that maximizes welfare with respect to the declared valuations and never charges agents more than their bids. Additionally, we consider other classes of valuations and bid spaces beyond those satisfying the gross substitutes conditions. Finally, we relax the no-overbidding assumption, and present bounds that are parameterized by the extent to which agents are willing to overbid.

3.3 The Tradeoff between Efficiency and Strategyproofness in Randomized Social Choice

Felix Brandt (TU München, DE)

License © Creative Commons BY 3.0 Unported license
© Felix Brandt

Joint work of Brandt, Felix; Aziz, Haris; Brandl, Florian; Brill, Markus

Main reference H. Aziz, F. Brandt, M. Brill, “On the Tradeoff Between Economic Efficiency and Strategy Proofness in Randomized Social Choice,” in Proc. of the 2013 Int’l Conf. on Autonomous Agents and Multi-agent Systems (AAMAS’13), pp. 455–462, IFAAMAS, 2013.

URL <http://dl.acm.org/citation.cfm?id=2484993>

Two fundamental notions in microeconomic theory are efficiency—no agent can be made better off without making another one worse off—and strategyproofness—no agent can obtain a more preferred outcome by misrepresenting his preferences. When social outcomes are probability distributions (or lotteries) over alternatives, there are varying degrees of these notions depending on how preferences over alternatives are extended to preference over lotteries. We show that efficiency and strategyproofness are incompatible to some extent when preferences are defined using stochastic dominance (SD) and therefore introduce a natural weakening of SD based on Savage’s sure-thing principle (ST). While random serial dictatorship is SD-strategyproof, it only satisfies ST-efficiency. Our main result is that strict maximal lotteries—an appealing class of social decision schemes due to Kreweras and Fishburn—satisfy SD-efficiency and ST-strategyproofness.

3.4 Algorithms for Strategic Agents I: Revenue Maximization

Yang Cai (MIT, US)

License © Creative Commons BY 3.0 Unported license
© Yang Cai

Joint work of Cai, Yang; Daskalakis, Constantinos; Weinberg, S. Matthew

Main reference Y. Cai, C. Daskalakis, S. M. Weinberg, “Optimal Multi-Dimensional Mechanism Design: Reducing Revenue to Welfare Maximization,” arXiv:1207.5518v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1207.5518v1>

In his seminal paper, Myerson [1981] provides a revenue-optimal auction for a seller who is looking to sell a single item to multiple bidders. Extending this auction to simultaneously selling multiple heterogeneous items has been one of the central problems in Mathematical Economics. We provide such an extension that is also computationally efficient. Our solution proposes a novel framework for mechanism design by reducing mechanism design problems (where one optimizes an objective function on “rational inputs”) to algorithm design problems (where one optimizes an objective function on “honest inputs”). Our reduction is generic and provides a framework for many other mechanism design problems.

3.5 Economic Efficiency Requires Interaction

Shahar Dobzinski (Weizmann Institute – Rehovot, IL)

License © Creative Commons BY 3.0 Unported license
© Shahar Dobzinski

Joint work of Dobzinski, Shahar; Nisan, Noam; Oren, Sigal

Main reference S. Dobzinski, N. Nisan, S. Oren, “Economic Efficiency Requires Interaction,” arXiv:1311.4721v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1311.4721v1>

We study the necessity of interaction between individuals for obtaining approximately efficient allocations. The role of interaction in markets has received significant attention in economic thinking, e.g. in Hayeks 1945 classic paper. We consider this problem in the framework of simultaneous communication complexity. We analyze the amount of simultaneous communication required for achieving an approximately efficient allocation. In particular, we consider two settings: combinatorial auctions with unit demand bidders (bipartite matching) and combinatorial auctions with subadditive bidders. For both settings we first show that non-interactive systems have enormous communication costs relative to interactive ones. On the other hand, we show that limited interaction enables us to find approximately efficient allocations.

3.6 Towards More Practical Linear Programming-based Techniques for Algorithmic Mechanism Design

Khaled Elbassioni (Masdar Institute – Abu Dhabi, AE)

License © Creative Commons BY 3.0 Unported license
© Khaled Elbassioni

Joint work of Elbassioni, Khaled; Ramezani, Fahimeh

Techniques based on linear programming, e.g., VCG-mechanism for fractional output sets and its extension to discrete output sets (Lavi and Swamy, 2005) for designing truthful(-in-expectation) mechanisms can be applied to many problems including combinatorial auctions. However, a direct implementation of these methods would be highly inefficient in practice, due to their reliance on general LP solvers, such as the Ellipsoid method. We investigate the possibility of using the much simpler and usually faster multiplicative weights update methods from convex optimization to speed-up these VCG- based techniques.

3.7 A unified approach to restricted complements

Michal Feldman (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Michal Feldman

Joint work of Feige, Uriel; Feldman, Michal; Immorlica, Nicole; Izsak, Rani; Lucier, Brendan; Syrgkanis, Vasilis

We study the efficiency of simultaneous single-item auctions when bidders have valuations that include restricted complementarities between items. We introduce and analyze a class of valuations that generalizes various notions of restricted complementarities, including supermodular degree recently introduced by Feige and Izsak (ECCC 2013), hypergraph valuations (Abraham et al. EC 2012), as well as monotone graphical valuations with positive

and negative weights (Acemoglu et al. 2012). For the introduced class of valuations, we show that the price of anarchy of simultaneous first-price item auctions is at most $2k$, where k describes the degree of complementarity. Our analysis proceeds via the smoothness framework, and therefore also applies to Bayesian equilibria and learning outcomes. Finally, we extend our results to the simultaneous composition of smooth mechanisms (e.g. simultaneous position auctions) under valuations that allow for restricted complements across mechanisms. One implication of this extension is that the price of anarchy of simultaneous second-price auctions is at most 2, for general bidder valuations, under a standard no-overbidding assumption.

3.8 Online price of anarchy for parking

Amos Fiat (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Amos Fiat

Joint work of Ardenboim, Alon; Cohen, Ilan; Colini-Baldeschi, Riccardo; Fiat, Amos

We show almost tight upper and lower bounds on the price of anarchy for parking in an unweighted line graph *sqrtn* and for arbitrary graphs.

3.9 Optimal Impartial Selection

Felix Fischer (University of Cambridge, GB)

License © Creative Commons BY 3.0 Unported license
© Felix Fischer

Joint work of Fischer, Felix; Klimm, Max

Main reference F. Fischer, M. Klimm, “Optimal Impartial Selection,” arXiv:1310.8631v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1310.8631v1>

We study the problem of selecting a member of a set of agents based on impartial nominations by agents from that set. The problem was studied previously by Alon et al. (TARK, 2011) and Holzman and Moulin (Econometrica, 2013) and has applications in situations where representatives are selected from within a group or where publishing or funding decisions are made based on a process of peer review. Our main result concerns a randomized mechanism that in expectation selects an agent with at least half the maximum number of nominations. Subject to impartiality, this is best possible.

3.10 Manipulation of Stable Matchings using Minimal Blacklists

Yannai A. Gonczarowski (The Hebrew Univ. of Jerusalem, IL)

License © Creative Commons BY 3.0 Unported license
© Yannai A. Gonczarowski

Main reference Y. A. Gonczarowski, “Manipulation of Stable Matchings using Minimal Blacklists,” arXiv:1307.7477v3 [cs.GT], 2013.

URL <http://arxiv.org/abs/1307.7477v3>

Gale and Sotomayor (1985) have shown that in the Gale-Shapley matching algorithm (1962), the proposed-to side W (referred to as women there) can strategically force the W -optimal stable matching as the M -optimal one by truncating their preference lists, each woman

possibly blacklisting all but one man. As Gusfield and Irving have already noted in 1989, no results are known regarding achieving this feat by means other than such preference-list truncation, i.e. by also permuting preference lists.

We answer Gusfield and Irving’s open question by providing tight upper bounds on the amount of blacklists and their combined size, that are required by the women to force a given matching as the M -optimal stable matching, or, more generally, as the unique stable matching. Our results show that the coalition of all women can strategically force any matching as the unique stable matching, using preference lists in which at most half of the women have nonempty blacklists, and in which the average blacklist size is less than 1. This allows the women to manipulate the market in a manner that is far more inconspicuous, in a sense, than previously realized. When there are less women than men, we show that in the absence of blacklists for men, the women can force any matching as the unique stable matching without blacklisting anyone, while when there are more women than men, each to-be-unmatched woman may have to blacklist as many as all men. Together, these results shed light on the question of how much, if at all, do given preferences for one side a priori impose limitations on the set of stable matchings under various conditions. All of these results are constructive, providing efficient algorithms for calculating the desired strategies.

3.11 Optimal Competitive Auctions

Nick Gravin (Microsoft Research New England – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Nick Gravin

Joint work of Chen, Ning; Gravin, Nick; Lu, Pinyan

Main reference N. Chen, N. Gravin, P. Lu, “Optimal Competitive Auctions,” arXiv:1401.0880v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1401.0880v1>

We study the design of truthful auctions for selling identical items in unlimited supply (e.g., digital goods) to n unit demand bidders. This classic problem stands out from profit-maximizing auction design literature as it requires no probabilistic assumptions on the buyers and employs the framework of competitive analysis. Our objective is to optimize the worst-case performance of an auction, measured by the ratio between a given benchmark and revenue generated by the auction.

We establish a sufficient and necessary condition that characterizes competitive ratios for all monotone benchmarks. The characterization identifies the worst-case distribution of instances and reveals intrinsic relations between competitive ratios and benchmarks in the competitive analysis. With the characterization at hand, we show optimal competitive auctions for two natural benchmarks.

The most well-studied benchmark measures the envy-free optimal revenue where at least two buyers win. Goldberg et al. (2004) have a sequence of lower bounds on the competitive ratio for each number of bidders n . They conjectured that all these bounds are tight. We show that optimal competitive auctions match these bounds. We confirm their conjecture and settle a central open problem in the design of digital goods auctions. As one more application we examine another economically meaningful benchmark, which measures the optimal revenue across all limited-supply Vickrey auctions. We identify optimal competitive ratios to be $(1 + 1/(n - 1))^{(n-1)-1}$ for each number of buyers n , that is $e - 1$ as n goes to infinity.

3.12 Quantitative Comparative Statics for a Multimarket Paradox

Tobias Harks (Maastricht University, NL)

License © Creative Commons BY 3.0 Unported license
© Tobias Harks

Joint work of Harks, Tobias; von Falkenhausen, Philipp

Main reference T. Harks, P. von Falkenhausen, “Quantitative Comparative Statics for a Multimarket Paradox,” arXiv:1307.5617v3 [cs.GT], 2013.

URL <http://arxiv.org/abs/1307.5617v3>

Comparative statics is a well established research field where one analyzes how changes in parameters of a strategic game affect the resulting equilibria. Examples of such parameter changes include tax or subsidy changes in oligopoly models or trade changes. While classic comparative statics is mainly concerned with qualitative approaches (e.g., deciding whether a parameter change improves or hurts equilibrium profits or welfare), we aim at quantifying this effect. We consider the famous multimarket oligopoly model introduced by Bulow, Geanakoplos and Klemperer. In this model, there are two firms competing on two markets with one firm having a monopoly on one market. Bulow et al. describe the counterintuitive example of a positive price shock in the firm’s monopoly market resulting in a reduction of the firm’s new equilibrium profit. We quantify for the first time the worst-case profit reduction for the case of two markets with affine price functions and firms with convex cost technologies. We show that the relative loss of the monopoly player is at most 25% no matter how many firms compete on the second market. In particular we show for the setting of Bulow et al. involving affine price functions and only one additional firm on the second market that the worst case loss in profit is bounded by 6.25%. We further investigate a dual effect: How much can a firm gain from a negative price shock in its monopoly market? Our results imply that this gain is at most 33%. We complement our bounds by concrete examples of markets where these bounds are attained.

3.13 Redesigning the Israeli Psychology Market

Avinatan Hassidim (Google Israel – Tel-Aviv, IL)

License © Creative Commons BY 3.0 Unported license
© Avinatan Hassidim

Joint work of Hassidim, Avinatan; Rom, Assaf

We present theoretical and practical issues that arose in the redesign process of the Israeli psychology market.

3.14 Designing Profit Shares in Coalition Formation Games

Martin Hoefer (Universität des Saarlandes, DE)

License © Creative Commons BY 3.0 Unported license
© Martin Hoefer

Joint work of Hoefer, Martin; Wagner, Lisa

Main reference M. Hoefer, L. Wagner, “Designing Profit Shares in Matching and Coalition Formation Games,” in Proc. of the 9th Int’l Conf. on Web and Internet Economics (WINE’13), LNCS, Vol. 8289, pp. 249–262, Springer, 2013.

URL http://dx.doi.org/10.1007/978-3-642-45046-4_21

Matching and coalition formation are fundamental problems in many scenarios where agents join efforts to perform tasks, such as, e.g., in scientific publishing. To allocate credit or

profit stemming from a joint project, different communities use different crediting schemes in practice. A prominent approach is equal sharing, where every member receives the same credit for a joint work. It captures a natural egalitarian fairness condition when each member of a coalition is critical for success. Unfortunately, when coalitions are formed by rational agents, equal sharing can lead to high inefficiency of the resulting stable states. We study how to design profit shares to obtain good stable states in matching and coalition formation games. We relax equal sharing to sharing schemes where for each coalition each player is guaranteed to receive at least an α -share. Using such schemes we characterize the tension between efficiency and equal treatment, and provide polynomial-time algorithms for their computation.

3.15 Privacy-Preserving Auctions

Zhiyi Huang (Stanford University, US)

License © Creative Commons BY 3.0 Unported license
© Zhiyi Huang

Joint work of Hsu, Justin; Huang, Zhiyi; Roth, Aaron; Roughgarden, Tim; Wu, Zhiwei Steven

Main reference J. Hsu, Z. Huang, A. Roth, T. Roughgarden, Z. S. Wu, “Private Matchings and Allocations,” arXiv:1311.2828v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1311.2828v1>

We consider a private variant of the classical allocation problem: given m goods and n agents with individual, private valuation functions over bundles of goods, how can we partition the goods amongst the agents to maximize social welfare? Specifically, the valuation functions are sensitive information which the agents wish to keep private from arbitrary coalitions of other agents. An important special case is when each agent desires at most one good, and specifies her (private) value for each good: in this case, the problem is exactly the maximum-weight matching problem in a bipartite graph.

Private matching and allocation problems have not been considered in the differential privacy literature, and for good reason: they are plainly impossible to solve under the standard notion of differential privacy. Informally, the allocation must match agents to preferred goods in order to maximize social welfare, but this preference is exactly what agents wish to keep private! Therefore, we consider the problem under the recently introduced constraint of joint differential privacy: roughly, for any agent i , no coalition of agents excluding i should be able to learn about the valuation function of agent i . We first show that if there are a small number of identical copies of each good, then it is possible to efficiently and accurately solve the maximum weight matching problem while guaranteeing joint differential privacy. We then extend our techniques to the more general allocation problem, when bidder valuations satisfy the gross substitutes condition. Finally, we prove lower bounds demonstrating that the problem cannot be privately solved to non-trivial accuracy without requiring multiple copies of each type of good.

3.16 Duality and optimality of auctions for the uniform distribution

Elias Koutsoupias (University of Oxford, GB)

License © Creative Commons BY 3.0 Unported license
© Elias Koutsoupias

We derive exact optimal solutions for the problem of optimizing revenue in single-bidder multi-items auctions for i.i.d. uniform distribution valuations. We give optimal auctions of

up to 6 items; previous results were only known for two items. To do so, we develop a general duality framework for the problem of maximizing revenue in many-bidders multi-item additive Bayesian auctions. The framework extends linear programming duality and complementarity to constraints with partial derivatives. The dual system reveals the geometric nature of the problem and highlights its connection with the theory of bipartite graph matchings. It is used both for deriving the optimal auction, which happens to be deterministic, and for proving optimality.

3.17 Characterization of SMON mechanisms with additive valuations over the real domain

Annamaria Kovacs (*Goethe-Universität Frankfurt am Main, DE*)

License © Creative Commons BY 3.0 Unported license
© Annamaria Kovacs

Joint work of Christodoulou, Giorgos; Kovacs, Annamaria; Vidali, Angelina

We are interested in the limits of characterizability of mechanisms with multi-dimensional, additive player-valuations like unrelated scheduling or additive combinatorial auctions. We characterize decisive, strongly monotone mechanisms for two tasks or items as either task independent mechanisms or '(player-)grouping minimizer's, a generalization of affine minimizers. (Further assumptions are the continuity of the payment functions, and that the bids are arbitrary real values.) This is work in progress: we strongly conjecture that the results generalize to m tasks/items by inductive arguments. We present a general lemma implying the linearity of payment functions in regular cases.

3.18 Mechanisms for Multi-Unit Combinatorial Auctions with a Few Distinct Goods

Piotr Krysta (*University of Liverpool, GB*)

License © Creative Commons BY 3.0 Unported license
© Piotr Krysta

Joint work of Krysta, Piotr; Telelis, Orestis; Ventre, Carmine

Main reference P. Krysta, O. Telelis, C. Ventre, "Mechanisms for Multi-unit Combinatorial Auctions with a Few Distinct Goods," in Proc. of the 2013 Int'l Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS'13), pp. 691–698, IFAAMAS, 2013.

URL <http://dl.acm.org/citation.cfm?id=2485029>

We design and analyze deterministic truthful approximation mechanisms for multi-unit combinatorial auctions with only a *constant* number of distinct goods, each in arbitrary limited supply. Prospective buyers (bidders) have preferences over *multisets* of items, i.e. for more than one unit per distinct good. Our objective is to determine allocations of multisets that maximize the Social Welfare. Despite the recent theoretical advances on the design of truthful combinatorial auctions (for several distinct goods) and multi-unit auctions (for a single good), results for the combined setting are much scarcer. Our main results are for *multi-minded* and *submodular* bidders. In the first setting each bidder has a positive value for being allocated one multiset from a prespecified *demand set* of alternatives. In the second setting each bidder is associated to a submodular valuation function that defines his value for the multiset he is allocated.

For multi-minded bidders we design a truthful FPTAS that *fully* optimizes the Social Welfare, while violating the supply constraints on goods within factor $(1 + \epsilon)$ for any fixed $\epsilon > 0$ (i.e., the approximation applies to the constraints and not to the social welfare). This result is best possible, in that full optimization is impossible without violating the supply constraints. It also improves significantly upon a related result of Grandoni *et al.* [SODA 2010]. For submodular bidders we extend a general technique by Dobzinski and Nisan [JAIR, 2010] for multi-unit auctions, to the case of multiple distinct goods. We use this extension to obtain a PTAS that approximates the optimum social welfare within factor $(1 + \epsilon)$ for any fixed $\epsilon > 0$, without violating the supply constraints. This result is best possible as well. Our allocation algorithms are *Maximum-in-Range* and yield truthful mechanisms when paired with Vickrey-Clarke-Groves payments.

3.19 Prior-free Auctions with Ordered Bidders

Stefano Leonardi (University of Rome “La Sapienza”, IT)

License © Creative Commons BY 3.0 Unported license
© Stefano Leonardi

Joint work of Bhattacharya, Sayan; Koutsoupias, Elias; Kulkarni, Janardhan; Leonardi, Stefano; Roughgarden, Tim; Xu, Xiaoming

Main reference S. Bhattacharya, E. Koutsoupias, J. Kulkarni, S. Leonardi, T. Roughgarden, X. Xu, “Near-optimal Multi-unit Auctions with Ordered Bidders,” in Proc. of the 14th ACM Conf. on Electronic Commerce (EC’13), pp. 91–102, ACM, 2013.

URL <http://doi.acm.org/10.1145/2482540.2482555>

Auctions are traditionally evaluated in economics theory using average-case or Bayesian analysis, and expected auction performance is optimized with respect to a prior distribution over inputs. Worst-case guarantees are desirable when, for example, good prior information is expensive or impossible to acquire, and when a single auction is to be re-used several times, in settings with different or not-yet-known input distributions. In this talk, we present prior-free auctions with constant-factor approximation guarantees in both unlimited and limited supply that also apply to a relevant case of non identical bidders. These auctions are simultaneously near-optimal in a wide range of Bayesian multi-unit environments when compared against the performance of Myerson optimal bayesian auction.

3.20 Implementing the “Wisdom of the Crowd”

Yishay Mansour (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Yishay Mansour

Joint work of Ilan, Kremer; Mansour, Yishay; Motty, Perry

Main reference I. Kremer, Y. Mansour, M. Perry, “Implementing the “Wisdom of the Crowd”,” in Proc. of the 14th ACM Conf. on Electronic Commerce (EC’13), pp. 605–606, ACM, 2013.

URL <http://doi.acm.org/10.1145/2482540.2482542>

We study a novel mechanism design model in which agents each arrive sequentially and choose one action from a set of actions with unknown rewards. The information revealed by the principal affects the incentives of the agents to explore and generate new information. We characterize the optimal disclosure policy of a planner whose goal is to maximize social welfare. One interpretation of our result is the implementation of what is known as the “wisdom of the crowd”. This topic has become increasingly relevant with the rapid spread of the Internet over the past decade.

3.21 Deferred Acceptance Auctions

Paul Milgrom (Stanford University, US)

License  Creative Commons BY 3.0 Unported license
© Paul Milgrom

Joint work of Milgrom, Paul; Segal, Ilya

We study auctions in which allocations are decided by an iterative process of rejecting the least attractive remaining bids. These *deferred-acceptance heuristic auctions* have distinctive properties that make them attractive for applications in computationally challenging environments. Deferred acceptance *threshold* auctions are group strategy-proof, can be implemented using clock auctions, and are outcome-equivalent in our complete-information model to paid-as-bid auctions based on the same heuristic. Paid-as-bid auctions based on such heuristics are dominance solvable, and every non-bossy dominance-solvable paid-as-bid auction is a deferred-acceptance heuristic auction. None of these properties are shared by auctions based on optimization or greedy-acceptance heuristics.

3.22 Plasticity, Monotonicity and Implementability

Rudolf Mueller (Maastricht University, NL)

License  Creative Commons BY 3.0 Unported license
© Rudolf Mueller

Joint work of Carbajal, Juan Carlos; Mueller, Rudolf

Consider a setting in which agents have quasilinear utilities over money and social alternatives. The set of alternatives can be finite or infinite. A domain D of admissible valuation functions of an agent is called a 2-cycle (3-cycle) monotonicity domain if every 2-cycle (3-cycle) monotone allocation rule defined on D is truthfully implementable in dominant strategies. It is called a revenue equivalence domain if every truthfully implementable allocation rule defined on D satisfies the revenue equivalence property. We introduce the notions of weak and strong plasticity, and prove that (i) every weak plasticity domain is a 3-cycle monotonicity and revenue equivalence domain; and (ii) very strong plasticity domain is a 2-cycle monotonicity and revenue equivalence domain. Our proof is elementary and does not rely on strenuous additional machinery. We also show various economic environments, with countable or uncountable allocations, in which weak and strong plasticity are satisfied.

3.23 Revenue Maximization with Sampling

Noam Nisan (The Hebrew University of Jerusalem, IL)

License  Creative Commons BY 3.0 Unported license
© Noam Nisan

Recent work in revenue-maximizing mechanism design has pursued, from an algorithmic perspective, multi-parameter extensions to Myerson's optimal single item auction. Much progress has been made, though much of it makes strong assumptions on the representation and/or structure of distributions from which players' values are drawn. We examine the single-buyer unit-demand mechanism design problem in its most general form, where the buyers' value distribution is presented as a "black box." We seek to understand the extent to

which revenue-maximizing mechanism design is possible in this general setting, and begin an exploration of the description complexity, sample complexity, and computational complexity of approximately revenue-maximizing auctions in the black-box model.

3.24 Dynamic Models of Reputation and Competition in Job-Market Matching

Sigal Oren (The Hebrew University of Jerusalem, IL)

License  Creative Commons BY 3.0 Unported license
© Sigal Oren

Joint work of Kleinberg, Jon; Oren, Sigal

A fundamental decision faced by a firm hiring employees – and a familiar one to anyone who has dealt with the academic job market, for example – is deciding what caliber of candidates to pursue. Should the firm try to increase its reputation by making offers to higher-quality candidates, despite the risk that the candidates might reject the offers and leave the firm empty-handed? Or is it better to play it safe and go for weaker candidates who are more likely to accept the offer? The question acquires an added level of complexity once we take into account the effect one hiring cycle has on the next: hiring better employees in the current cycle increases the firm’s reputation, which in turn increases its attractiveness for higher-quality candidates in the next hiring cycle. These considerations introduce an interesting temporal dynamic aspect to the rich line of research on matching models for job markets, in which long-range planning and evolving reputational effects enter into the strategic decisions made by competing firms.

We develop a model that captures these effects in a setting where two firms repeatedly compete for job candidates over multiple periods. Within this model, we attempt to estimate the effect that reasoning about future hiring cycles has on the efficiency of the job market: do people end up unnecessarily unemployed while the firms compete over the top candidates, or does the evolution of reputation over time eventually converge to a two-tiered system in which the firms each target different parts of the market?

3.25 Matchings, Vertex Cover und Network Bargaining Games

Britta Peis (RWTH Aachen, DE)

License  Creative Commons BY 3.0 Unported license
© Britta Peis

In an instance of the classical, cooperative *matching game* introduced by Shapley and Shubik [Int. J. Game Theory '71] we are given an undirected graph $G = (V, E)$, and we define the value $\nu(S)$ of each subset $S \subseteq V$ as the cardinality of a maximum matching in the subgraph $G[S]$ induced by S . The *core* of such a game contains all *fair* allocations of $\nu(V)$ among the players of V , and is well-known to be non-empty iff graph G is *stable*. G is stable if its inessential vertices (those that are exposed by at least one maximum matching) form a stable set.

In this paper we study the following natural edge-deletion question: given a graph $G = (V, E)$, can we find a minimum-cardinality *stabilizer*? I.e., can we find a set F of edges whose removal from G yields a stable graph?

We show that this problem is vertex-cover hard. We then prove that there is a minimum-cardinality stabilizer that avoids some maximum-matching of G . We employ this insight to give efficient approximation algorithms for sparse graphs, and for regular graphs.

3.26 Learning Equilibria of Games via Payoff Queries

Rahul Savani (University of Liverpool, GB)

License © Creative Commons BY 3.0 Unported license
© Rahul Savani

Joint work of Fearnley, John; Gairing, Martin; Goldberg, Paul; Savani, Rahul

Main reference J. Fearnley, M. Gairing, P. Goldberg, R. Savani, “Learning Equilibria of Games via Payoff Queries,” arXiv:1302.3116v3 [cs.GT], 2013.

URL <http://arxiv.org/abs/1302.3116v3>

We study a computational learning model for games in which an algorithm queries the payoffs of players at pure strategy profiles. The goal of the algorithm is to find an exact or approximate Nash equilibrium of the game with as few queries as possible. We give basic results on the payoff query complexity of bimatrix and graphical games. We then focus on symmetric network congestion games. For directed acyclic networks, we can learn the cost functions (and hence compute an equilibrium) while querying just a small fraction of pure-strategy profiles. For the special case of parallel links, we have the stronger result that an equilibrium can be identified while only learning a small fraction of the cost values.

3.27 Non Adaptive Methods for Adaptive Seeding

Yaron Singer (Harvard University, US)

License © Creative Commons BY 3.0 Unported license
© Yaron Singer

Joint work of Badanidiyuru, Ashwin; Lattanzi, Silvio; Papadimitriou, Christos; Rubinfeld, Aviad; Seaman, Lior

Adaptive seeding is a two-stage stochastic optimization framework recently developed for information dissemination in social networks. The goal is to optimize a combinatorial function by making an initial decision that affects the realizations selected by nature. Beyond information dissemination in networks other interesting applications are in machine learning and operations research. In this talk we will discuss several optimization techniques for adaptive seeding as well as results in social network analysis that motivate this approach.

3.28 Cost-Recovering Bayesian Algorithmic Mechanism Design

Balasubramanian Sivan (Microsoft Research – Redmond, US)

License © Creative Commons BY 3.0 Unported license
© Balasubramanian Sivan

Joint work of Fu, Hu; Lucier, Brendan; Sivan, Balasubramanian; Syrgkanis Vasilis

Main reference H. Fu, B. Lucier, B. Sivan, V. Syrgkanis, “Cost-Recovering Bayesian Algorithmic Mechanism Design,” arXiv:1305.0598v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1305.0598v1>

Consider a group of participants competing to receive service from a mechanism that can provide such services at a cost. The mechanism aims to serve agents to maximize social

efficiency, without suffering an expected loss: the agent’s payments should cover the service cost in expectation. We develop a general method for converting arbitrary approximation algorithms for the underlying optimization problem into Bayesian incentive compatible mechanisms that are cost-recovering in expectation.

3.29 Composable and Efficient Mechanisms

Éva Tardos (Cornell University, US)

License © Creative Commons BY 3.0 Unported license
© Éva Tardos

Joint work of Syrgkanis, Vasilis; Tardos, Éva

Main reference V. Syrgkanis, É. Tardos, “Composable and Efficient Mechanisms,” arXiv:1211.1325v1 [cs.GT], 2012.

URL <http://arxiv.org/abs/1211.1325v1>

In this talk, we consider auctions as games, and we discuss how to analyze such games providing robust guarantees for their performance even when players participate in multiple auctions, have valuations that are complex functions of multiple outcomes, and are using learning strategies to deal with an uncertain environment.

3.30 An Optimal Online Algorithm for Weighted Bipartite Matching and Extensions to Packing Linear Programs

Andreas Toennis (RWTH Aachen, DE)

License © Creative Commons BY 3.0 Unported license
© Andreas Toennis

Joint work of Thomas, Kesselheim; Klaus, Radke; Tönnis, Andreas; Berthold, Vöcking

Main reference T. Kesselheim, K. Radke, A. Tönnis, B. Vöcking, “An Optimal Online Algorithm for Weighted Bipartite Matching and Extensions to Combinatorial Auctions,” in Proc. of the 21st Annual European Symp. on Algorithms (ESA’13), LNCS, Vol. 8125, pp. 589–600, Springer, 2013.

URL http://dx.doi.org/10.1007/978-3-642-40450-4_50

We present an ϵ -competitive algorithm for online weighted bipartite matching in the random order model. In this model a bipartite edge-weighted graph is given by an adversary. The vertices on the right-hand side are given in advance, while the left-hand side vertices arrive online in a random order. Whenever a vertex arrives his adjacent edges with the corresponding weights are revealed and the online algorithm has to decide which of these edges should be included in the matching.

Furthermore we extend the approach to packing linear programs. Here the capacity vector is given in advance and columns, thus variables, arrive in a random order. With every variable, its contribution to the target function and its consumption of resources is revealed. In this setting we also provide an optimal algorithm that is $1 - O(\sqrt{(1 + \log d)/\epsilon^2})$ -competitive where d is the maximal number of non-zero entries in a column. This algorithm can be turned into a truthful mechanism using VCG payments. Additionally the algorithm is not based on a primal-dual approach but solely depends on the primal solution and therefore it can be combined with any approximation algorithm.

3.31 Local computation mechanism design

Shai Vardi (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Shai Vardi

Joint work of Hassidim, Avinatan; Mansour, Yishay; Vardi, Shai

Main reference A. Hassidim, Y. Mansour, S. Vardi, “Local computation mechanism design,” arXiv:1311.3939v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1311.3939v1>

We introduce the notion of Local Computation Mechanism Design – designing game theoretic mechanisms which run in polylogarithmic time and space. Local computation mechanisms reply to each query in polylogarithmic time and space, and the replies to different queries are consistent with the same global feasible solution. In addition, the computation of the payments is also done in polylogarithmic time and space. Furthermore, the mechanisms need to maintain incentive compatibility with respect to the allocation and payments.

We present local computation mechanisms for a variety of classical game- theoretical problems: (1) stable matching, (2) job scheduling, (3) combinatorial auctions for unit-demand and k -minded bidders, and (4) the housing allocation problem.

For stable matching, some of our techniques may have general implications. Specifically, we show that when the men’s preference lists are bounded, we can achieve an arbitrarily good approximation to the stable matching within a fixed number of iterations of the Gale-Shapley algorithm.

3.32 Algorithms for Strategic Agents II

S. Matthew Weinberg (MIT, US)

License © Creative Commons BY 3.0 Unported license
© S. Matthew Weinberg

Joint work of Cai, Yang; Daskalakis, Constantinos; Weinberg, S. Matthew

Main reference Y. Cai, C. Daskalakis, S. M. Weinberg, “Understanding Incentives: Mechanism Design becomes Algorithm Design,” arXiv:1305.4002v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1305.4002v1>

We provide a computationally efficient black-box reduction from mechanism design to algorithm design. Specifically, we give an approximation-preserving reduction from truthfully optimizing any objective with arbitrary bidder types to algorithmically optimizing the same objective plus virtual welfare. Furthermore, we extend the reduction to accommodate a bi-criterion approximation algorithm that we call (α, β) -approximations. We apply our framework to obtain the following results:

1. This reduction is tight for revenue. That is, we also give an approximation-sensitive reduction from optimizing virtual welfare algorithmically to optimizing revenue truthfully.
2. As an application of 1), it is NP-hard to approximately maximize revenue for a single monotone submodular bidder within any poly($\#$ items) factor.
3. A 10.5-approximate truthful mechanism for minimizing makespan on unrelated machines.

3.33 Utility-Target Auctions

Christopher A. Wilkens (Yahoo Labs – Sunnyvale, US)

License © Creative Commons BY 3.0 Unported license
© Christopher A. Wilkens

Joint work of Hoy, Darrell; Jain, Kamal; Wilkens, Christopher A.

Main reference D. Hoy, K. Jain, C. A. Wilkens, “A Dynamic Axiomatic Approach to First-Price Auctions,” arXiv:1304.7718v1 [cs.GT], 2013.

URL <http://arxiv.org/abs/1304.7718v1>

The first-price auction is popular in practice for its simplicity and transparency. Moreover, its potential virtues grow in complex settings where incentive compatible auctions may generate little or no revenue. Unfortunately, generalizing the first-price auction has proven fragile in theory and practice.

We show that the auctioneer’s choice of bidding language is critical when generalizing beyond the single-item setting, and we propose a specific construction called the utility-target auction that performs well. The utility-target auction includes a bidder’s final utility as an additional parameter, identifying the single dimension along which she wishes to compete. This auction is closely related to profit-target bidding in first-price and ascending proxy package auctions and gives strong performance guarantees for a variety of complex auction environments.

We also take a dynamic approach to studying pay-your-bid auctions: rather than basing performance guarantees solely on static equilibria, we study the repeated setting and show that robust performance guarantees may be derived from simple axioms of bidder behavior. For example, as long as a loser raises her bid quickly, a standard first-price auction will generate at least as much revenue as a second-price auction. We generalize such ideas to complex pay-your-bid auctions through the utility-target auction: as long as losers do not wait too long to raise bids, a first-price auction will reach an envy-free state that implies a strong lower-bound on revenue; as long as winners occasionally experiment by lowering their bids, the outcome will near the boundary of this envy-free set so bidders do not overpay; and when players with the largest payoffs are the least patient, bids converge to the egalitarian equilibrium. Significantly, bidders need only know whether they are winning or losing in order to implement such behavior.

3.34 A Unified Approach to Truthful Scheduling on Related Machines

Rob van Stee (University of Leicester, GB)

License © Creative Commons BY 3.0 Unported license
© Rob van Stee

Joint work of Epstein, Leah; Levin, Asaf; van Stee, Rob

Main reference L. Epstein, A. Levin, R. van Stee, “A unified approach to truthful scheduling on related machines,” arXiv:1207.3523v1 [cs.DS], 2012.

URL <http://arxiv.org/abs/1207.3523v1>

We present a unified framework for designing deterministic monotone polynomial time approximation schemes (PTAS’s) for a wide class of scheduling problems on uniformly related machines. This class includes (among others) minimizing the makespan, maximizing the minimum load, and minimizing the p -norm of the machine loads vector. Previously, this kind of result was only known for the makespan objective. Monotone algorithms have the property that an increase in the speed of a machine cannot decrease the amount of work assigned to it. The key idea of our novel method is to show that for goal functions that

are sufficiently well-behaved functions of the machine loads, it is possible to compute in polynomial time a highly structured nearly optimal schedule. An interesting aspect of our approach is that, in contrast to all known approximation schemes, we avoid rounding any job sizes or speeds throughout. We can therefore find the exact best structured schedule using dynamic programming. The state space encodes a sufficient amount of information such that no postprocessing is needed, allowing an elegant and relatively simple analysis without any special cases. The monotonicity is a consequence of the fact that we find the best schedule in a specific collection of schedules. Monotone approximation schemes have an important role in the emerging area of algorithmic mechanism design. In the game-theoretical setting of these scheduling problems there is a social goal, which is one of the objective functions that we study. Each machine is controlled by a selfish single-parameter agent, where its private information is its cost of processing a unit sized job, which is also the inverse of the speed of its machine. Each agent wishes to maximize its own profit, defined as the payment it receives from the mechanism minus its cost for processing all jobs assigned to it, and places a bid which corresponds to its private information. For each one of the problems, we show that we can calculate payments that guarantee truthfulness in an efficient manner. Thus, there exists a dominant strategy where agents report their true speeds, and we show the existence of a truthful mechanism which can be implemented in polynomial time, where the social goal is approximated within a factor of $1 + \epsilon$ for every $\epsilon > 0$.

Participants

- Yossi Azar
Tel Aviv University, IL
- Moshe Babaioff
Microsoft Research –
Mountain View, US
- Felix Brandt
TU München, DE
- Yang Cai
MIT, US
- Shahar Dobzinski
Weizmann Institute –
Rehovot, IL
- Khaled Elbassioni
Masdar Institute –
Abu Dhabi, AE
- Michal Feldman
Tel Aviv University, IL
- Amos Fiat
Tel Aviv University, IL
- Felix Fischer
University of Cambridge, GB
- Yannai A. Gonczarowski
The Hebrew University of
Jerusalem, IL & Microsoft
Research – Redmond, US
- Nick Gravin
Microsoft Research –
Cambridge, US
- Tobias Harks
Maastricht University, NL
- Paul Harrenstein
University of Oxford, GB
- Avinatan Hassidim
Google Israel – Tel-Aviv, IL
- Monika Henzinger
Universität Wien, AT
- Martin Hoefer
Universität des Saarlandes, DE
- Zhiyi Huang
Stanford University, US
- Elias Koutsoupias
University of Oxford, GB
- Annamaria Kovacs
Goethe-Universität Frankfurt am
Main, DE
- Piotr Krysta
University of Liverpool, GB
- Stefano Leonardi
University of Rome “La
Sapienza”, IT
- Yishay Mansour
Tel Aviv University, IL
- Paul Milgrom
Stanford University, US
- Rudolf Müller
Maastricht University, NL
- Noam Nisan
The Hebrew University of
Jerusalem, IL
- Sigal Oren
The Hebrew University of
Jerusalem, IL
- Britta Peis
RWTH Aachen, DE
- Heiko Röglin
Universität Bonn, DE
- Rahul Savani
University of Liverpool, GB
- Yaron Singer
Harvard University, US
- Balasubramanian Sivan
Microsoft Research –
Redmond, US
- Alexander Skopalik
Universität Paderborn, DE
- Éva Tardos
Cornell University, US
- Andreas Tönnis
RWTH Aachen, DE
- Rob van Stee
University of Leicester, GB
- Shai Vardi
Tel Aviv University, IL
- Bernhard von Stengel
London School of Economics, GB
- S. Matthew Weinberg
MIT, US
- Christopher A. Wilkens
Yahoo Labs – Sunnyvale, US



Computational Models of Language Meaning in Context

Edited by

Hans Kamp¹, Alessandro Lenci², and James Pustejovsky³

1 Universität Stuttgart, DE, Hans.Kamp@ims.uni-stuttgart.de

2 University of Pisa, IT, alessandro.lenci@ling.unipi.it

3 Brandeis University, Waltham, USA, jamesp@cs.brandeis.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 13462 “Computational Models of Language Meaning in Context”. The seminar addresses one of the most significant issues to arise in contemporary formal and computational models of language and inference: that of the role and expressiveness of distributional models of semantics and statistically derived models of language and linguistic behavior. The availability of very large corpora has brought about a near revolution in computational linguistics and language modeling, including machine translation, information extraction, and question-answering. Several new models of language meaning are emerging that provide potential formal interpretations of linguistic patterns emerging from these distributional datasets. But whether such systems can provide avenues for formal and robust inference and reasoning is very much still uncertain. This seminar examines the relationship between classical models of language meaning and distributional models, and the role of corpora, annotations, and the distributional models derived over these data. To our knowledge, there have been no recent Dagstuhl Seminars on this or related topics.

Seminar 10.–15. November, 2013 – www.dagstuhl.de/13462

1998 ACM Subject Classification 1.2.7 Natural Language Processing

Keywords and phrases formal semantics, distributional semantics, polysemy, inference, compositionality, Natural Language Processing, meaning in context

Digital Object Identifier 10.4230/DagRep.3.11.79

1 Executive Summary

Hans Kamp

Alessandro Lenci

James Pustejovsky

License  Creative Commons BY 3.0 Unported license
© Hans Kamp, Alessandro Lenci, and James Pustejovsky

The term distributional semantics qualifies a rich family of computational methods sharing the assumption that the statistical distribution of words in context plays a key role in characterizing their semantic behavior. Distributional semantic models, such as LSA, HAL, etc., represent the meaning of a content word in terms of a distributed vector recording its pattern of co-occurrences (sometimes, in specific syntactic relations) with other content words within a corpus. Different types of semantic tasks and phenomena are then modeled in terms of linear algebra operations on distributional vectors. Distributional semantic models provide a quantitative correlate to the notion of semantic similarity, and are able to address various lexical semantic tasks, such as synonym identification, semantic classification, selectional preference modeling, and so forth.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Computational Models of Language Meaning in Context, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 79–116

Editors: Hans Kamp, Alessandro Lenci, and James Pustejovsky



DAGSTUHL
REPORTS

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Distributional semantics has become increasingly popular in Natural Language Processing. Its attractiveness lies in the fact that distributional representations do not require manual supervision and reduce the the a priori stipulations in semantic modeling. Moreover, distributional models generally outperform other types of formal lexical representations, such as for instance semantic networks. Many researchers have also strongly argued for the psychological validity of distributional semantic representations. Corpus-derived measures of semantic similarity have been assessed in a variety of psychological tasks ranging from similarity judgments to simulations of semantic and associative priming, showing a high correlation with human behavioral data.

Despite its successes, no single distributional semantic model meets all requirements posed by formal semantics or linguistic theory, nor do they cater for all aspects of meaning that are important to philosophers or cognitive scientists. In fact, the distributional paradigm raises the question of the extent to which semantic properties can be reduced to combinatorial relations. Many central aspects of natural language semantics are left out of the picture in distributional semantics, such as predication, compositionality, lexical inferences, quantification and anaphora, just to quote a few. A central question about distributional models is whether and how distributional vectors can also be used in the compositional construction of meaning for constituents larger than words, and ultimately for sentences or discourses – the traditional domains of denotation-based formal semantics. Being able to model key aspects of semantic composition and associated semantic entailments represents a crucial condition for distributional model to provide a more general model of meaning. Conversely, we may wonder whether distributional representations can help to model those aspects of meaning that notoriously challenge semantic compositionality, such as semantic context-sensitivity, polysemy, predicate coercion, pragmatically-induced reference and presupposition.

The main question is whether the current limits of distributional semantics represent contingent shortcomings of existing models – hopefully to be overcome by future research –, or instead they point to intrinsic inadequacies of vector-based representations to address key aspects of natural language semantics. To this end, there were five themes addressed by the participants:

1. The problems in conventional semantic models that distributional semantics claims to be able to solve;
2. The promise of distributional semantics linking to multimodal representations
3. The current limitations of distributional semantics theories to account for linguistic compositionality;
4. The absence of any robust first-order models of inference for distributional semantics;
5. The integration of distributional semantic principles and techniques into a broader dynamic model theoretic framework.

2 Table of Contents

Executive Summary

Hans Kamp, Alessandro Lenci, and James Pustejovsky 79

Overview of Talks

Towards a distributionally motivated formal semantics of natural language

Hans Kamp 83

Model Theory and Distributional Semantics

Katrin Erk 83

Implicative uses of evaluative factive adjectives

Lauri Karttunen 83

Formal Semantics and Distributional Semantics: A Survey of Chance and Challenge

Hinrich Schütze 84

Working Groups 84

Statements on Distributional Semantics by the Seminar Participants 85

Dr Strangestats or How I Learned to Stop Worrying and Love Distributional Semantics

Marco Baroni 85

Position statement

Peter Cariani 86

Position statement

Stephen Clark 90

Position statement

Ann Copestake 93

Position statement

Katrin Erk 94

Position statement

Stefan Evert 97

Statements on Distributional Semantics

Sebastian Löbner 98

Putting together the pieces

Louise McNally 98

Incrementality in Compositional Distributional Semantics

Alessandra Zarcone, Sebastian Padó 101

Acquiring conceptual knowledge for semantic interpretation

Massimo Poesio 102

Research overview

Tim Van de Cruys 105

Semantics, Communication, and Probability

Jan van Eijck 107

Position statement <i>Dominic Widdows</i>	108
Norms and Exploitations in Text Meaning and Word Use <i>Patrick Hanks</i>	109
Position statement <i>Anna Rumshisky</i>	110
Position statement <i>Annie Zaenen</i>	112
Panel Discussions	113
Polysemy	113
Inference	113
Compositionality	114
Negation	114
Next Steps	115
Participants	116

3 Overview of Talks

3.1 Towards a distributionally motivated formal semantics of natural language

Hans Kamp (Universität Stuttgart, DE)

License  Creative Commons BY 3.0 Unported license
© Hans Kamp

Formal models of semantics for natural language have proved to be very powerful and useful in their description of linguistic phenomena. To date, there is no distributional semantic model that satisfies the requirements posed by formal semantics or linguistic theory for modeling meaning. Nor does distributional semantics to my knowledge address issues of meaning that are important to philosophers or cognitive scientists. In fact, the distributional paradigm raises the question of the extent to which semantic properties can be reduced to combinatorial relations. Many central aspects of natural language semantics, such as predication, compositionality, lexical inferences, quantification and anaphora, seem left out of the picture. The challenge is to find models that have the explanatory adequacy of formal semantic theories, but which are at the same time able to capture the contextual and distributional nature of language use.

3.2 Model Theory and Distributional Semantics

Katrin Erk (University of Texas – Austin, US)

License  Creative Commons BY 3.0 Unported license
© Katrin Erk

What is the denotation of distributional representations? It seems reasonable to say that purely linguistic data can change our beliefs about the world. But does this also hold for distributional information? After all, distributional data just counts sentential contexts in which words have been observed, and it is not clear how we could derive truth conditions from a distributional vector. But distributional information can do something less: It can provide tentative, uncertain information about similarities between different predicates that have been mentioned in the text. And this, I think, suffices to reduce our uncertainty about which world we are in. We can describe this in a probabilistic semantics setting. We have a prior probability distribution over worlds, which can be updated (in a standard fashion, by Baye's rule) using uncertain distributional information.

3.3 Implicative uses of evaluative factive adjectives

Lauri Karttunen (Stanford University, US)

License  Creative Commons BY 3.0 Unported license
© Lauri Karttunen

Evaluative adjectives such as stupid, smart, lucky, sweet, cruel can take propositional complements as in *John was smart to leave early*. In this construction they are generally considered to be factive: *John was not smart to leave early* is supposed to mean that John did leave early and that it was not a good idea. When one looks at the web, however, one finds easily examples where the intended meaning is not factive, as in *Luckily, I was not*

stupid to send them any money. Amazon MT experiments and corpus studies confirm the existence of this pattern. We will discuss whether it should be seen as a performance error or whether there seems to be a real 'dialect' split among speakers of (American) English. Whatever the analysis, it seems that the pattern is prevalent enough for NLP applications that assign factuality judgments to events to need to take it into account.

3.4 Formal Semantics and Distributional Semantics: A Survey of Chance and Challenge

Hinrich Schütze (LMU München, DE)

License  Creative Commons BY 3.0 Unported license
© Hinrich Schütze

I will first describe what I see as the strengths and weaknesses of distributional models on the one hand and formal models on the other hand. I will then contrast two different types of distributional models, count vector models and deep learning embedding models. The main part of the talk will be about compositionality and about the extent to which distributional and formal semantic models can handle different aspects of compositionality.

4 Working Groups

Participants were assigned to one of four groups, each discussing a specific set of questions related to the seminar topic. The topics are given below.

1. **Polysemy and Vagueness**
 - type coercion, metonymy, complex types,
 - metaphor, figurative language
 - issues of lexical inference (for non-function words)
 - semantic relations
2. **Inference and Reasoning**
 - structural deduction based on the representational syntax, axioms, and inference rules.
 - Inference from a DS perspective: computation over and similarity of vectors?
 - What to do about Quantification
 - But inference is not just deduction; Can DS distinguish between deduction, induction, and abduction?
 - Defeasibility and default logics how do these stand up against the more natural soft constraints given by distributional techniques and probabilistic reasoning.
3. **Compositionality**
 - function application
 - selectional preferences are handled well in DS. What about type shifting?
 - semantic roles,
 - Basic semantics of predication in DS
4. **Modality and Negation**
 - Negation: difficult to handle in DS.
 - Tense: put it on the map
 - Deontic logic
 - Epistemic logic and reasoning about knowledge and beliefs

The group composition was as follows:

Group 1: Stefan Evert, Tim van de Cruys, Patrick Hanks, Sebastian Löbner, Suzanne Stevenson, Alessandra Zarcone

Group 2: Ann Copestake, Ido Dagan, Jan van Eijck, Graeme Hirst, Sebastian Padó, Anna Rumshisky, Dominic Widdows

Group 3: Marco Baroni, Stephen Clark, Katrin Erk, Jerry Hobbs, Alessandro Lenci, Louise McNally, Massimo Poesio

Group 4: Nicholas Asher, Peter Cariani, Hans Kamp, Lauri Karttunen, James Pustejovsky, Hinrich Schütze, Mark Steedman, Annie Zaenen

5 Statements on Distributional Semantics by the Seminar Participants

5.1 Dr Strangestats or How I Learned to Stop Worrying and Love Distributional Semantics

Marco Baroni (University of Trento, IT)

License  Creative Commons BY 3.0 Unported license
© Marco Baroni

I was a teenage generativist. I was raised in fairly observant Chomskyan schools, and I still abide by the program for linguistics as the algorithmic study of human language competence Chomsky laid out 60 years ago. Then, how did I become an adult werelinguist, theoretical semanticist by day, corpus-based, statistics-driven computationalist at night? I don't do corpus-based, statistics-driven distributional semantics because I am, in principle, attracted by or sympathetic towards usage-based, nonsymbolic, inductive approaches to language. I do distributional semantics because at a certain point I discovered that it is the only semantic formalism allowing me to do my job as a linguist. I first felt the need for semantics while writing my master thesis about derivational morphology, where the salience of morpheme boundaries predicts phenomena such as the likelihood that an affix undergoes phonetic reduction, blocking of phonological rules, morphemic-route access in lexical retrieval, etc. But one of the main factors determining, in turn, the salience of morpheme boundaries is semantic transparency, that is, the extent to which the meaning of a derived word is related to the meaning of its stem, (cf. re-decorate vs. recollect). I then started looking around for an approach to semantics that would (i) provide large-scale coverage of the lexicon and (ii) make quantitative predictions about degrees of similarity (or relatedness). The first requirement came from the fact that I needed to account for the often semantically arbitrary sets of stems and derived forms that were subject to specific morpheme salience phenomena. The second requirement derived from the fact that, in all phenomena I looked into, the effect of semantic transparency was never all or nothing, but rather a fuzzy phenomenon with many intermediate cases, so I needed a theory making graded predictions.

Formal approaches to semantics, even those that paid attention to lexical meaning, failed both requirements. The functionalist stuff, while in principle sympathetic to the idea of degrees of similarity, was too awfully fuzzy, not explicit enough to make quantitative predictions, and in any case failing the coverage requirement. Unfortunately, I discovered distributional semantics too late to use it in my morphology work, where I just gave up the idea of accounting for semantic transparency effects, but from when, years later, I discovered

LSA and its cousins, I never found a reason to go back to other approaches to semantics, simply because, from a practical point of view, I still see no alternative to the distributional approach. Something I've learned along the way is that being able to quantify degrees of semantic similarity is not only good for tasks such as assessing the semantic transparency of derived forms or finding near synonyms. Distributional semanticists (including some that will attend this seminar) came up with clever and elegant ideas to account, in terms of semantic similarity, for complex linguistic phenomena such as predicting the selectional preferences of verbs, capturing argument alternation classes or accounting for co-composition effects. And there is ongoing and very promising work (that, I think, will be discussed at the seminar) on dealing with fundamental challenges for distributional semantics such as polysemy or scaling up to phrase and sentence meaning. So, while there is a lot of hard work ahead of us, I'm confident that in a few years we will have empirically successful models of distributional semantics that are not limited to single words in isolation, and, equipped with these new models, we will be able to account for many more linguistic phenomena in terms of semantic similarity.

Still, current distributional semantics is entirely prisoned inside a linguistic cage: all it can tell us (and that's not little) is how similar words, phrases and sentences are to each other. Without a hook into the outside world, all we will be able to do is to measure how similar, say, the sentence "A boy is laughing" is to "A girl is crying", but we will never be able to tell whether either sentence can be truthfully asserted of the current state of the world. While I understand that there is much more to the outside world than this, I think that one first, reasonable step we can take is to explore whether we can connect distributional semantic representations with our visual perception of the world. In concrete, we should aim for a system that, given a picture depicting a scene with, say, a laughing boy, could tell us that A boy is laughing is an appropriate statement describing the scene. Interestingly, state-of-the-art image analysis systems represent images not unlike distributional semantics represents words – that is, images are represented by vectors that record the distribution of a set of discrete feature occurrences in them. So, there is hope, and I think a central goal for distributional semantics in the next few years should be to work on how to develop a common semantic space, where vector-based representations of linguistic expressions, on one side, and objects and scenes, on the other, can be mapped and compared. Given such shared linguistic-visual semantic space, the same similarity scoring techniques we are already using in distributional semantics might be extended to account for referential aspects of meaning: The sentence "A boy is laughing" is truthfully stated of (a picture depicting) a scene if the vector representing the sentence and the vector representing the scene are above a certain threshold of similarity. My colleagues and I are currently working on methods to build the proposed shared linguistic-visual semantic space (and other researchers are also making good progress in this direction). At the seminar, I would like to discuss (among many other things, of course) both concrete ideas about how to construct the common space, and what are linguistically interesting scenarios in which we could make use of it.

5.2 Position statement

Peter Cariani (Harvard Medical School – Newton, US)

License  Creative Commons BY 3.0 Unported license
© Peter Cariani

I think that perhaps I am the outlying point, the wild card here, so I will try to explain myself. I come to questions of meaning acquisition/construction from a naturalistic, pragmatic,

constructivist perspective that is heavily influenced by cybernetics and systems theory; auditory, computational, and theoretical neuroscience; and perceptual and cognitive psychology. I currently teach courses related to the neuropsychology of music, which I think has some deep parallels with the kinds of computational semantics questions we have here before us. I hope that it will be useful to the group have an independent external perspective. My intention is not to distract or detain you from the nuts-and-bolts aspects of the specific questions at hand before us (I'm sure you would all be happy just hashing through the minutiae of your sub-fields, as would I), but I want to try to stand back (since this is the only place I can stand here) and raise broader questions when (if) needed. Here are some of the basic issues I see:

- (I) Statistics vs. structure in interpretation/anticipation
- (II) Why do semantic analysis? Computational tools vs. modeling minds/brains
- (III) Getting pragmatics into computational semantics
- (IV) Implementations: Symbols vs. connectionism vs. something else
- (V) Can computational semantics (ultimately) understand human life?

I. Statistics vs. structure in interpretation/anticipation

A. Statistics-based learning and interpretation. The fundamental issues really go back to old and unresolved debates about the how minds and brains work, i.e. how much of human cognition is driven by the statistics of external input patterns vs. by the internal organization of mental processes. The answer is that both aspects play important roles, that minds/brains are anticipatory systems that register, remember, and act upon external event statistics, albeit very heavily filtered through a powerful mental apparatus that ever attempts to predict the future by constructing highly structured models of the world. What is (could be) the relationship between these two kinds of anticipatory processes, in minds, brains, and machines? For example in this current discourse before us, I see the distributional semantics approach as part of a larger resurgence of associationist psychology that I believe has been fueled in recent decades by the (perceived and real) successes of hidden-Markov models for automatic speech recognition. In the neurosciences, over the past two decades, there has been a blossoming of interest in Bayesian perceptual models and the statistics of natural scenes. These methods have their own practical applications and efficacies, but every powerful information technology eventually becomes a model of minds and brains for some fraction of those who use it. We need to be clear about whether our purpose is to develop computational tools that serve as adjuncts to our own reasoning and meaning-making (e.g. more effective search engines or corpora analyzers) or whether we are trying to model human mental processes of meaning formation. Coming out of the applied mathematics of statistics-based machine learning, it seems to me that distributional semantics tends to view itself as a set of useful techniques (and perhaps the mind as an assemblage of such hacks, as Minsky thinks). The underlying (often tacit) assumptions of these models are that minds (sensory, cognitive systems) do not have strong internal structure and adapt to the statistics of incoming information (in whatever modality or form).

B. Structure-based learning and interpretation. On the other hand, are what I think of as structuralist theories, in the old psychological sense of that term, a la Titchener and Piaget and the Gestaltists, that hold that there is strong dimensional structure to mental processes, and that therefore it is necessary to model those structural constraints if we are to understand and predict human interpretation and to replicate its functionalities in artificial systems. Logic- and model-based approaches to semantics share with structuralist psychology that there are strong constraints (I use the term low dimensional structure), and

that the crux of understanding systematicity and compositionality lies in the underlying sets of basic informational processes (symbols and rules, neural/mental representations and operations) that are operant in logics/models/minds/brains. Clearly both kinds of mechanisms are operant in minds and brains. There is widespread evidence that humans and animals learn the statistics of their surrounds and adapt to them, such that, in the absence of better predictive information, they will produce expectancies based on those statistics. For example, we see this in music perception when listeners are exposed to artificial scales (e.g. Bohlen–Pierce) and come to expect those musical intervals - they adapt to the pitch statistics of their recent experience. However, this statistical prediction is a weak expectancy, and it is easily superseded if there is strong predictive structure in the music (repeating phrases, motifs, sections, rhythmic patterns, etc.). Musical expectancy is a combination of what I call pattern (structure) and frame (statistics)³ I am currently working on neural timing net models for rhythmic pattern expectancy – when there is longer range repeating structure, that dominates; in lieu of longer-range structure, basic event probabilities dominate. In speech perception, I think we only use prior phoneme and word probabilities when signal-to-noise ratios are low – otherwise, when signals are clear, deterministic auditory pattern recognition processes dominate and we can easily achieve 100 accuracy identifying strings of nonsense syllables and words.

II. Developing effective computational tools vs. modeling the mind

I can see already that the different approaches (model-based vs. distributional semantics) have different purposes. The latter can leverage the awesome power of computer statistical analysis over extremely large and varied digital corpora. The former, however, hold out the even greater promise of an eventual theory of how minds make meanings, and if we can solve those hard problems, we shall have much more effective digital analysis technologies. In this discussion we need to be as clear as possible about what our goals are re: computational semantics – otherwise we will discuss and/or argue at cross-purposes.

III. Getting pragmatics into computational semantics

Pragmatic frames should be central to both model-theoretic and distributional semantics. The perceived intended purposes of communications we receive and texts that we interpret play heavily into how we interpret the meaning of the message. In terms of forming interpretive meanings of human and animal communications, I think pragmatics comes first, semantics second, and syntactics third. We humans are already primed heavily by the situational pragmatics to assume the nature of the message (neutral communication, threat, warning, command, question, affective expression, etc.), that in turn bias selection of semantic senses, that rapidly form a conceptual model of the contents of the message. We then do a detailed syntactic analysis if there are unresolved incongruities, if the model doesn't make any sense or if the contents of the message don't comport with the nature of the communication. This has practical implications. The distributional strategy of product (or music) recommendation based on co-occurrences of looks or purchases (those people who looked at this eventually bought that, or those people who like this music also liked that) is useful in that it indicates a correlation that may have a relevant underlying cause. However, the correlation does not inform the prospective buyer of why people liked or bought those things they chose. Really, except for purely imitative buyers/listeners, this is what we want to know, why we should choose one thing over another. We want to choose our music by what we want from it (e.g. happy/sad/interesting/comforting/surprising/nostalgic/arousing/sleep

inducing/meditative/distracting/identity-affirming/etc. music). Music recommendation systems are beginning to do this, but it requires a structural theory of the effects of different kinds of music and musical parameters on internal psychological states. Purposes of actors and pragmatic contexts can be incorporated into model-theoretic accounts (as pragmatist philosophy holds, truth is really efficacy relative to purpose). There are also ways that distributional approaches might incorporate pragmatic observables. I believe that brains encode information in a manner that allows a given event, object, or association to be content-addressable via all of its manifold aspects: pragmatic, semantic, syntactic. This is how we solve Dreyfus' frame problem, we can search by purpose, by effect on the world, by effect on us, by form and bring up those relevant dimensional aspects that we need in a given situation.

IV. Symbols vs. connectionism vs. something else

I am a theoretical-computational auditory neuroscientist and have been dealing with the whole issue of time codes in the brain. The focus of this workshop is not on neural models per se, but in essence I think that the distributional-model-theoretic semantics discussion has many parallels with the symbols vs. connectionism debate twenty years ago. I had a front row seat at the MIT debate between Smolensky, champion of connectionist neural nets, and the tag team Pylyshyn and Fodor, champions of symbolic computations. I was rooting for the neural networks, but symbols easily carried the day. How minds realize universals, abstract categories, systematicity, and compositionality are fundamental problems that neuroscience and psychology need to solve in order to construct an adequate theory of mind. It's a useful heuristic to try to imagine how minds work, i.e. how brains operate to form meanings and interpretations. In terms of neural activity patterns, it appears to me that all of these aspects simultaneously activate in parallel respective sets of neural assemblies (a la Lashley and Hebb) that in effect resonate with each other to different degrees, such that subsets of neural assemblies implementing different interpretations reinforce each other, with different pattern-resonant subsets competing with other subsets. The end result is a parallel-analysis and competitive winner-take-all process, but one in which later, conflicting information can reverse earlier dominant interpretations (defeasible constraints).

V. Can computational semantics (ultimately) understand what it means to be human?

I know this is very philosophical, but maybe it is worth thinking about. I think we should always try to think as far ahead as we can about where (how far) these theories can take us. Even above the questions of the respective efficacies and limitations of model-theoretic vs. distributional semantics that we will hash out here, there are some general questions of the extent to which formal systems (either logic-based models or full-blown psychological models of human minds) can capture private and public meanings. It would seem to me that if we had an adequate theory of the brain, such that we could simulate its information processing aspects properly, including sensorimotor transactions with the environment and embedded internal reward systems, that we could have an adequate model for human meaning. I don't believe that the brain or mental processes involved are necessarily logical in the truth-theoretic sense (e.g. each of us simultaneously holds sets of logically conflicting beliefs; moral and political reasoning is notoriously based on competing modes of thinking that are based on largely complementary imperatives). This begs the question of whether a computer, using only the encoded text resources of the internet, could possibly understand what it is like to be human and to interpret texts in those terms (in terms of the meanings of things).

Or in other words, would the machine need itself to have needs, drives, feelings, friends, enemies, memories to interpret texts in the ways that we do? This sounds like Hubert Dreyfus' frame problem (which I think brains solve by encoding all memories of events and their hedonic outcomes in pragmatic and semantic terms – we have memory that is content-addressable both by semantics–world effect and pragmatics–use effect. The midline dopamine systems encode the internal time– sequences of all the neural events that lead up to reward or punishment, such that both relations between perceived world–events and relations between actions, world–events, and rewards can be predicted). Can all those aspects of our internal structure and our interactions with the world be modeled and simulated, such that the machine will extract a meaning that is similar to one we would produce?

References

- 1 Cariani, P. 1989. On the Design of Devices with Emergent Semantic Functions. Ph.D., State University of New York at Binghamton.
- 2 Cariani, P. A. Delgutte, B. 1996a. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol*, 76, 1698–716. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J Neurophysiol*, 76, 1717–34.
- 3 Cariani, P. 2001a. Cybernetics and the semiotics of translation. In: Petrilli, S. (ed.) *Lo Stesso Altro: Athanor: arte, letteratura, semiotica, filosofia*, v. XII, n. 4. Rome: Meltemi.
- 4 Cariani, P. A. 2001b. Neural timing nets. *Neural Netw*, 14, 737–53.
- 5 Cariani, P. 2002. Temporal codes, timing nets, and music perception. *J. New Music Res.*, 30, 107–136.
- 6 Cariani, P. 2011. The semiotics of cybernetic percept-action systems. *International Journal of Signs and Semiotic Systems*, 1, 1–17.
- 7 Cariani, P. 2012. Creating new primitives in minds and machines. In: McCormack, J., D'Inverno, M. (eds.) *Computers and Creativity*. New York: Springer.
- 8 Cariani, P., Micheyl, C. 2012. Towards a theory of information processing in the auditory cortex. In: Poeppel, D., Overath, T., Popper, A. (eds.) *Human Auditory Cortex: Springer Handbook of Auditory Research*. New York: Springer.

5.3 Position statement

Stephen Clark (University of Cambridge, GB)

License  Creative Commons BY 3.0 Unported license
© Stephen Clark

My current research goal is to develop compositional techniques for distributional semantics. This goal is relevant for the scientific enterprise of computational linguistics, since the accounts of distributional semantics currently lack a satisfactory compositional treatment; and also for the engineering enterprise of natural language processing (NLP), since representing the meanings of phrases and larger units in a vector space will allow the calculation of semantic similarity for those phrases and larger units. Calculating semantic similarity is crucial for many NLP tasks and applications. In collaboration with Bob Coecke (Oxford) and Mehrnoosh Sadrzadeh (Queen Mary), I have developed a tensor-based theoretical framework for distributional semantics which applies readily to variants of categorial grammar [5]. The idea is that the syntactic type of a constituent determines its semantic type; for example, the meaning of a transitive verb in English is represented as a 3^{rd} order tensor. Tensors are multi-linear maps in multi-linear algebra; hence the framework encapsulates the old idea

from formal semantics that the meanings of some words and phrases can be represented as functions. The compositional operation which combines the tensors is tensor contraction (just matrix multiplication extended to the multi-linear algebra case). Moreover, since tensors are functions, the combinatory operations of Combinatory Categorical Grammar (CCG) [14] transfer over to the tensor-based framework in a straightforward way, meaning that the framework applies to CCG as well as the context-free pregroup grammars of Lambek used in our original papers [11].

A useful instance of the framework to consider is adjective-noun modification. Since the syntactic type of an adjective in English is N/N , its semantic type is $N \otimes N$; a vector in $N \otimes N$ is a matrix representing a function from the noun space onto the noun space. Composition of an adjective and noun is achieved with matrix multiplication. In fact this is the proposal of [2] (independently conceived), which can be seen as an instance of our more general framework. The framework is currently largely a theoretical framework, with some small-scale attempts at implementation [8, 7, 10], and additional theoretical work building on it [3]. There are a number of practical and theoretical stumbling blocks in the way of a large-scale implementation. Many of these stumbling blocks are fundamental questions relating to natural language semantics, and in particular semantics in context, and hence of relevance to the Seminar.

Questions

What is the sentence space? The theoretical framework assumes a separate vector space for sentences, S , compared with nouns, which live in N . (There may be other spaces corresponding to the basic syntactic types, also, for example PP .) However, the framework only dictates how to compose functions and arguments to deliver a vector in that (assumed) space; it does not place any constraints on what the sentence space should be. This raises the question of whether it makes sense to represent the meanings of sentences in a vector space, and how structured should such a sentence space be? The answer may depend on the application; for example for sentiment analysis, a simple space of positive/negative may suffice. Another way to ask the same question is whether phrases and sentences should live in the same space as nouns (as they do in the neural-network based work of Socher, for example [13]). Making this assumption simplifies the implementation, but it is questionable whether the semantics of sentences can be fully captured in a vector space designed to represent the semantics of nouns.

Should the composed representations be distributional? I make a distinction between a distributed representation – which I take to mean simply vector- (or tensor-) based – and a distributional representation, which I take to mean a representation based on contextual information (as in the classic vector-based representations of word meanings [12, 9]). [1] take the intriguing position that all distributed representations are distributional, including those at the phrase and sentence level. Another alternative is to suppose that the word representations – especially those of nouns – are distributional, but the representations of larger phrases are distributed, without necessarily reflecting the distributional contexts of those phrases in some large, idealized corpus.

Can higher-order tensors be built in practice? Whilst there are machine learning techniques in place for learning higher-order tensors, given some suitable objective function, in practice the task of learning tensors for all word-category pairs in the lexicon is a formidable one. Dimensionality reduction techniques may help, but it is likely that the order of some of the tensors will need reducing in the grammar. For example, syntactic types such as

$((N/N)/(N/N))/((N/N)/(N/N))$ are not uncommon in the output of the CC CCG parser [4], but this would result in an 8-order tensor, with a huge number of parameters.

Do we need both operator and contextual semantics The meaning of a transitive verb, for example, in the framework is a 3rd-order tensor (a function). This is what I am calling operator semantics. But of course a (1st-order) vector can also be built for transitive verbs, in the standard way (which I am calling contextual semantics). We can also build (distributed) representations of the selection preferences of the verb. Are these separate from the operator semantics? Do we need all these representations? Another way to consider this question is whether the proposal of Erk and Pado [6] would benefit from the addition of operator semantics as provided by our compositional tensor-based framework. One area where this question arises is in relative clauses. Here, a verb phrase, which has operator semantics in the framework (represented by a matrix), needs to combine, via the relative pronoun, with the noun, which has contextual semantics (represented by a vector) [3]. Hence there appears to be a typemismatch here. Providing an additional representation for the verb phrase – either its contextual vector, or its selectional preferences – and allowing that to combine with the noun (eg through pointwise multiplication) may solve this problem.

Can logical operators be incorporated into the framework? This question relates to the more general question of whether traditional notions from formal semantics – which could also include quantification and inference – can be incorporated into a vector-space setting. My work is currently less focused on this question, but it is obviously important. A more general question is whether formal semantics is needed in addition to distributional semantics, or whether there is an all-encompassing framework.

References

- 1 M. Baroni, R. Bernardi, and R. Zamparelli. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies*, 2013.
- 2 M. Baroni and R. Zamparelli. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the EMNLP Conference*, pages 1183–1193, Boston, MA, 2010.
- 3 Stephen Clark, Bob Coecke, and Mehrnoosh Sadrzadeh. The Frobenius anatomy of relative pronouns. In *Proceedings of the ACL Mathematics of Language workshop*, Sofia, Bulgaria, 2013.
- 4 Stephen Clark and James R. Curran. Widecoverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics*, 33(4):493–552, 2007.
- 5 Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis 36: A Festschrift for Joachim (Jim) Lambek*, 2010.
- 6 Katrin Erk and Sebastian Pado. A structured vector space model for word meaning in context. In *Proceedings of the EMNLP Conference*, Hawaii, 2008.
- 7 Edward Grefenstette and Mehrnoosh Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, 2011.
- 8 Edward Grefenstette, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke, and Stephen Pulman. Concrete sentence spaces for compositional distributional models of meaning. In *Proceedings of the 9th International Conference on Computational Semantics (IWCS-11)*, pages 125–134, Oxford, UK, 2011.
- 9 Gregory Grefenstette. *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers, 1994.

- 10 Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Stephen Pulman. A unified sentence space for categorical distributional-compositional semantics: Theory and experiments. In Proceedings of the International Conference on Computational Linguistics, Mumbai, India, 2012.
- 11 Joachim Lambek. *From Word to Sentence. A Computational Algebraic Approach to Grammar*. Polimetrica, 2008.
- 12 Hinrich Schütze. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123, 1998.
- 13 Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. Semantic compositionality through recursive matrix-vector spaces. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 1201–1211, Jeju, Korea, 2012.
- 14 Mark Steedman. *The Syntactic Process*. The MIT Press, Cambridge, MA, 2000.

5.4 Position statement

Ann Copestake (University of Cambridge, GB)

License  Creative Commons BY 3.0 Unported license
© Ann Copestake

The following brief (and unavoidably rushed) notes are partly drawn from the ‘Lexicalised compositionality’ draft paper available from my web page (joint work with Aurélie Herbelot). See also my position paper in the distributional semantics workshop in IWCS 2013.

Theoretical perspective. Distributional semantics is best seen as belonging to usage-based accounts of language. While the philosophical tradition is in many ways difficult and not very helpful as a guide for computational linguists (later Wittgenstein etc), some more modern work seems to provide a better basis: I find Brandom’s approach in ‘Making it explicit’ particularly helpful. Human languages can be used to ‘do logic’, but it doesn’t follow that that’s all language semantics is about. The notion of an ‘Ideal distribution’ in our lexicalised compositionality paper is an attempt to show under what conditions there is a relationship between a distributional account and a model-theoretic account. There is no reason why a notion of an individual (linguistic and/or real world) can’t be combined with a distributional account. This seems essential for modelling quantification, and (probably) also some lexical semantic phenomena such as antonymy. The role of generalization (inheritance) in distributions seems a promising area of investigation: the difference in this regard from previous approaches to lexical semantics is very striking.

Compositional semantics and distributions. I believe it is better to base distributions on a lightweight model of compositional semantics, such as (D)MRS, than on syntax, since there are cases of syntax-semantics mismatches (expletive ‘it’ etc, etc) which compositional semantics is well-equipped to deal with. Similarly, I currently see no need for distributional semantics to redo compositional accounts of tense (e.g., the English auxiliary system). Lightweight models allow for different interpretations of e.g., adjective noun combination, which gives scope for distributional semantics. Distributional semantics is particularly good at semi-compositional situations (cf derivational morphology).

Corpora. If we want to take the idea of psycholinguistic plausibility seriously (and I see this as a major advantage of distributional approaches), then we should work with realistic corpora. Collecting corpora based on an individual’s language experience should be a priority. Failing good notions of situated discourse in corpora, it may make sense to work with corpora

that exemplify one particular language game (or small class of language games) such as Wikipedia.

5.5 Position statement

Katrin Erk (University of Texas – Austin, US)

License  Creative Commons BY 3.0 Unported license
© Katrin Erk

Own previous work: Graded representations for word meaning in context. The topic that first got me interested in distributional models was the problem of representing word senses. Manually annotating documents with dictionary-based word senses is notoriously difficult, and both cognitive linguists [27, 7, 13] and lexicographers [15, 14] have cast doubt on the existence of clear-cut sense boundaries. Distributional models can be used to represent word meaning in context without reference to dictionaries if we compute a separate vector for each occurrence. Then we just have different occurrences that are closer together or further apart in space, without the need to draw sense boundaries. We have proposed a number of models for computing such occurrence representations [10, 11, 20].

Own previous work: Combining logic and distributional semantics. Distributional models have proved incredibly useful at the level of words and of short phrases. So what should be their role in sentence meaning representations? One possibility would be to use compositional distributional approaches to derive vectors for arbitrary sentences. But my hunch is that these vectors will become more and more noisy as phrase length and phrase complexity rise, where by noisy I mean that quite different phrases would receive similar vectors. (Table 1 in [24] seems to hint at something like this.) This is also my answer to Q3, the question about current limitations of compositional distributional semantics. It is my impression that the largest limitation of compositional distributional semantics lies in phrase length and complexity. Instead of pursuing a compositional distributional approach to sentence meaning, we are representing sentence meaning through logical form and are adding distributional similarity information (at both the word and short phrase level) as weighted inference rules [12, 4]. We use Markov Logic Networks [23] to do probabilistic inference on the resulting weighted clause set.

Q1: What can distributional semantics do that conventional semantic models cannot?

To me, the central reason to adopt distributional semantics is gradience, for example the ability to model degrees of similarity in word meanings [10, 26, 8, 28, 22]. A related strength of distributional models is their ability to describe relations between words through an open-ended list of possible phrases rather than through a fixed list of possible relations. Lapata and Lascarides [16] use this idea for logical metonymy. In a corpus-based model, the most likely interpretations for “begin song” that they derive are “sing”, “rehearse”, “write”, “hum”, “play”. Butnariu and Veale [5] use a similar idea for interpreting noun-noun compounds. Another big advantage of distributional models is coverage. They can extract usage-based information (representing a mixture of semantic and pragmatic phenomena) automatically for large numbers of lexical items.

Problems of distributional data. That said, I would like to list some problems of distributional semantics. They have all been discussed before, but I think they still need to be mentioned. The first problem is that we only have a single signal, co-occurrence, caused by a mixture of phenomena. A verb-noun cooccurrence can indicate a selectional preference

or an idiom [2]. Distributional similarity links near-synonyms (cup-mug) and pragmatically connected words (cup-milk) [21]. The second problem is lack of reference: We only have co-occurrence between words, no link between words and objects in the world [3]. The third problem is a reporting bias. Newspaper text tend to report on man bites dog, but not dog bites man [25].

Q5: Distributional semantics and model theory. I think that the time has come to revisit deep semantic analysis in computational linguistics. I have noticed more and more papers that talk about the need to address phenomena that deep semantic analysis is good at, like negation, modals, and implicatives. This happens in particular in textual entailment [1, 19, 18], but also in sentiment analysis [6]. And while it has been stated repeatedly that logic is “brittle”, I think it is not the logic that is brittle, but the inference mechanism and the background information available to the system. One way to address this is to use probabilistic inference, and to add distributional information.

There are currently two main approaches to combining distributional semantics and model theory. We transform distributional similarity to weighted distributional inference rules, and use probabilistic inference. Lewis and Steedman [17], on the other hand, use clustering on distributional data to infer word senses, but use standard first-order inference on the resulting logical forms. The main difference between the two approaches lies in the role of gradience. Lewis and Steedman view weights and probabilities as a problem to be avoided. We believe that the uncertainty inherent in both language processing and world knowledge should be front and center in the inference we do. (Though it is true that probabilistic inference is currently slow and memory-intensive.)

But in both current approaches to integrating distributional information into model-theoretic semantics, one important question is still open: What is the denotation of distributional representations? I have proposed interpreting distributional representations over conceptual structures [9], but that cannot be quite right: Given that distributional data is collected from texts of many speakers, it is not clear whose concepts these are supposed to be.

Here is a new proposal. It seems reasonable to say that purely linguistic data can change our beliefs about the world – that is what language does. But does this also hold for distributional information? After all, distributional data just counts sentential contexts in which words have been observed [3], and it is not clear how we could derive truth conditions from a distributional vector [29]. But distributional information can do something less: It can provide tentative, uncertain information about similarities between different predicates that have been mentioned in the text. And this, I think, suffices to reduce our uncertainty about which world we are in. We can describe this in a probabilistic semantics setting. We have a prior probability distribution over worlds, which can be updated (in a standard fashion, by Bayes’ rule) using uncertain distributional information.

References

- 1 Roy Bar-Haim, Ido Dagan, Iddo Grental, and Eyal Shnarch. Semantic inference at the lexical-syntactic level. In Proceedings of AAAI, Vancouver, Canada, 2007.
- 2 Marco Baroni and Alessandro Lenci. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721, 2010.
- 3 Marco Baroni, Raffaella Bernardi, and Roberto Zamparelli. Frege in space: A program for compositional distributional semantics. To appear.
- 4 Islam Beltagy, Cuong Chau, Gemma Boleda, Dan Garrette, Katrin Erk, and Raymond Mooney. Montague meets Markov: Deep semantics with probabilistic logical form. In Proceedings of *SEM, Atlanta, Georgia, USA, 2013.

- 5 Cristina Butnariu and Tony Veale. A concept-centered approach to nouncompound interpretation. In Proceedings of COLING, 2008.
- 6 Yejin Choi and Claire Cardie. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In Proceedings of EMNLP, Honolulu, Hawaii, 2008.
- 7 D. A. Cruse. Polysemy and related phenomena from a cognitive linguistic viewpoint. In P. Saint-Dizier and E. Viegas, editors, *Computational Lexical Semantics*, pages 33–49. Cambridge University Press, 1995.
- 8 Georgiana Dinu and Mirella Lapata. Measuring distributional similarity in context. In Proceedings of EMNLP, Cambridge, MA, 2010.
- 9 Katrin Erk. Towards a semantics for distributional representations. In Proceedings of IWCS, Potsdam, Germany, 2013.
- 10 Katrin Erk and Sebastian Padó. A structured vector space model for word meaning in context. In Proceedings of EMNLP, Honolulu, HI, 2008.
- 11 Katrin Erk and Sebastian Padó. Exemplar-based models for word meaning in context. In Proceedings of ACL, Uppsala, Sweden, 2010.
- 12 Dan Garrette, Katrin Erk, and Raymond Mooney. Integrating logical representations with probabilistic information using Markov logic. In Proceedings of IWCS, Oxford, UK, 2011.
- 13 Dirk Geeraerts. Vagueness’s puzzles, polysemy’s vagaries. *Cognitive linguistics*, 4:223–272, March 1993.
- 14 Patrick Hanks. Do word meanings exist? *Computers and the Humanities*, 34 (1-2):205–215(11), 2000.
- 15 Adam Kilgarriff. I don’t believe in word senses. *Computers and the Humanities*, 31(2):91–113, 1997.
- 16 Mirella Lapata and Alex Lascarides. A probabilistic account of logical metonymy. *Computational Linguistics*, 29(2):261–315, 2003.
- 17 Mike Lewis and Mark Steedman. Combined distributional and logical semantics. *TACL*, 1:179–192, 2013.
- 18 Amnon Lotan, Asher Stern, and Ido Dagan. Truthteller: Annotating predicate truth. In Proceedings of NAACL, 2013.
- 19 Bill MacCartney and Christopher D Manning. An extended model of natural logic. In Proceedings of IWCS, 2009.
- 20 Taesun Moon and Katrin Erk. An inference-based model of word meaning in context as a paraphrase distribution. *ACM Transactions on Intelligent Systems and Technology special issue on paraphrasing*, 2013.
- 21 Gregory L. Murphy. *The Big Book of Concepts*. MIT Press, 2002.
- 22 Diarmuid Ó Séaghdha and Anna Korhonen. Probabilistic models of similarity in syntactic context. In Proceedings of EMNLP, Edinburgh, UK, 2011.
- 23 Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine Learning*, 62:107–136, 2006.
- 24 Richard Socher, Eric H. Huang, Jeffrey Pennin, Andrew Y. Ng, and Christopher D. Manning. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In NIPS. 2011.
- 25 Mohammad S Sorower, Janardhan R Doppa, Walker Orr, Prasad Tadepalli, Thomas G Dietterich, and Xiaoli Z Fern. Inverting grice’s maxims to learn rules from natural language extractions. In NIPS, 2011.
- 26 Stefan Thater, Hagen Fürstenau, and Manfred Pinkal. Contextualizing semantic representations using syntactically enriched vector models. In Proceedings of ACL, Uppsala, Sweden, 2010.
- 27 David H. Tuggy. Ambiguity, polysemy and vagueness. *Cognitive linguistics*, 4 (2):273–290, 1993.

- 28 Tim Van de Cruys, Thierry Poibeau, and Anna Korhonen. Latent vector weighting for word meaning in context. In Proceedings of EMNLP, Edinburgh, Scotland, UK., 2011.
- 29 Jan van Eijck and Shalom Lappin. Probabilistic semantics for natural language. Draft published on the NASSLLI 2012 webpage, 2011.

5.6 Position statement

Stefan Evert (Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 3.0 Unported license
© Stefan Evert

At the current time, I am mainly interested in certain mathematical and practical issues of traditional distributional semantic models (DSMs) – which compile representations for single linguistic units (usually words) – and in practical applications that require broad – coverage semantics.

Research questions:

- Impact of DSM type (term–context vs. term–term) and model parameters (span–based vs. dependency–based co–occurrence, frequency weighting, normalization, dimensionality reduction, distance measure,) on the semantic representation.
- To what extent can DSMs be optimized for a particular task? Is there a single representation that captures general word meaning and works well for a broad range of tasks?
- Dimensionality reduction
 - Is it useful?
 - What are the differences between available methods (PCA, SVD, randomized SVD, RI, NMF, LDA,)?
 - How many latent dimensions should be used?
 - Are the reduced representations compatible with simple approaches to compositionality and polysemy?
- Ambiguity and polysemy: DSM vectors represent a weighted average over all senses of the corresponding word.
 - How can different meanings be identified and separated?
 - How can the context–dependent meaning of a word be computed?

If such “traditional” DSMs are applied to larger units (word pairs, phrases or sentences) these are either treated as opaque multiword units or a simplistic approximation to the compositional meaning is used (addition = Schütze’s bag–of–words approach, pointwise multiplication, etc.). I am interested in extensions of these approaches that capture some facets of compositionality while maintaining the simple structure and broad coverage of the simple models. Research questions include:

- Should compositional DSMs aim for a distributional representation (i.e. an approximation of the DSM vector of the phrase as an opaque unit) or merely a distributed representation (i.e. any vector representation; a typical example are convolution–based models such as HRR and BEAGLE).
- What are desirable properties for the DSM distances between compositional units and between such units and individual words?
 - Which of these properties are satisfied by addition, pointwise multiplication and other simple composition operations?

- Can simple compositional representations of large units (e.g. a bag-of-words model for sentences) be seen as a case of ambiguity/polysemy (e.g. a weighted mixture of topics present in a sentence)?

My research interests thus connect to, and overlap with themes 1 (polysemy and vagueness) and 3 (compositionality) of the seminar.

5.7 Statements on Distributional Semantics

Sebastian Löbner (Heinrich-Heine-Universität Düsseldorf, DE)

License  Creative Commons BY 3.0 Unported license
© Sebastian Löbner

General judgment. I consider Distributional Semantics a very important methodological achievement for the working semanticist as it offers tools and means for retrieving evidence and data for semantic analysis that were hitherto not available. It appears particularly promising for lexical decomposition, as the co-occurrence of lexical items provides evidence for their combinatory propensities and these, in turn, and to some degree, may provide access to their semantic content. Combinatory propensities are also relevant for a theory of composition, in particular by providing more data that will enable us to broaden the scope of theoretical analysis; but I think the qualitative rule-based approaches to a theory of syntactic, and concomitant semantic, composition are, and will keep being, superior to whatever can be gained by merely statistical methods.

Skepticism. I am very skeptical as to the potential of the DS approach for bringing us closer to an understanding of the cognition of language. Obviously, the brain does not work with this kind of software (the relevance of statistical weights for cognitive learning notwithstanding). As a theoretical semanticist, I am ultimately aiming at an understanding of the cognitive level of language. At present, it appears, there are encouraging developments in the cognitive sciences that open ways for developing cognitive semantic theories of decomposition and composition that take us crucially beyond the first cognitive approaches from the late 20th century (e.g. prototype theory) and the (indirect) insights into semantic cognition that were gained in formal semantics by logical analysis. Trying to model (?) or just do semantic composition with statistical methods might work one day to a certain degree of efficiency (similar to parsing, or machine translation by statistical methods) – but it will not bring us further to an understanding of semantic cognition.

5.8 Putting together the pieces

Louise McNally (UPF – Barcelona, ES)

License  Creative Commons BY 3.0 Unported license
© Louise McNally

I got into linguistics through cognitive science. I was interested in all kinds of big questions about language and the mind, but the more I thought about these questions the more I realized that I understood almost nothing about language, and I didn't think it made much sense to try to answer them without a better idea of how language works. So I decided to

go to grad school in linguistics, hoping that would give me the background I would need to continue in cognitive science. Little did I realize how complex language would turn out to be.

My main goal as a linguist is to understand how lexical meaning (if we can distinguish such a thing) is integrated with general conceptual knowledge, on the one hand, and information coming from reference, on the other, when we interpret phrases and produce them for others. I have pursued this goal by carrying out detailed studies of linguistic phenomena that cannot be understood without some theory of how this integration works. The two most relevant empirical areas I have worked on are modification – the construction of complex descriptive contents – and the many manifestations of the type (conceptual)/token (referential) distinction in language. I have developed most of my work using the tools associated with what I'll loosely refer to as the formal semantics community.

Though it is unquestionable that this community's focus on understanding the connection between words and the world, and on the development of the corresponding tools to do so, has led to huge advances in the theory of meaning, the limitations of this focus have long been known to everyone. One of the things I have found most unsatisfying about formal semantics (though not unsatisfying enough to give up on the whole enterprise) is that these limitations have mostly been quietly ignored. One extremely negative effect of this is that formal semantic research has arguably not had the impact on cognitive science that it could have, and my impression from the last IWCS is that its early contributions to computational linguistics are running a certain risk of being lost. On the bright side, various lines of formal and computational research have addressed several of these limitations. The big pending task is to bring these lines together in a systematic way. This seminar looks like a good opportunity to make some progress. In the rest of this statement, I briefly mention some of the issues that I think should be placed on the table for discussion.

Conceptual vs. referential aspects of meaning: Perhaps the most serious and long-standing problem in study of meaning is the division between approaches and, correspondingly, communities of researchers, according to whether the conceptual or referential dimension of meaning (to say nothing of the social) is the primary focus of interest. Discourse Representation Theory (in a particularly clear way in comparison to other dynamic logics) was perhaps the first systematic attempt to distinguish formally between reference and the descriptive conditions that the referents in our discourse models must satisfy. Though these conditions have generally been modeled as grounded in the world, I do not see any reason in principle why they could not be associated with conceptual contents. I think there is great potential here that is only beginning to be explored (see for example recent work by Erk and colleagues, Kamp, and myself with Baroni and Boleda in this direction).

Though a differentiated treatment of reference vs. descriptive content conditions is not the focus of the richly typed systems that e.g. Pustejovsky and Asher have used to develop more sophisticated analyses of the composition of lexical meanings, these approaches are certainly compatible with such a treatment. In contrast, it is far less obvious how to capture such a distinction in distributional models of meaning, or whether we should even try to do so. The implications of this characteristic of distributional models are profound; I'll come back to them briefly below.

The representation and composition of lexical meanings: The development of rich systems of types and type composition operations has made it possible to express important generalizations concerning the ways lexical meanings are typically modulated in the context of other lexical items – for instance, we can easily capture the role of part-whole relations or the function an entity typically has in accounting for patterns of metonymy. Distributional

semantic models improve in some ways on these systems, but at least in their present state arguably lose ground on others. Since distributional representations reflect not only lexical entailments but also a lot of other contextual information associated with an expression, they are very suggestive as a means of approximating richer conceptual representations, and their behavior under composition offers the hope of analyses of polysemy resolution and phenomena such as metaphor that are more general and finer-grained than those afforded by symbolic systems. However, it is less clear how the composition of distributional representations can be modulated to reflect the salience of the sorts of relations embodied in e.g. qualia structures and that arguably have psychological reality independently of the sheer frequency of their occurrence.

Function words, content words, and the syntax/semantics/discourse interface: We commonly distinguish between so-called content words and function words, the latter serving, for example, to help manage referential relations (e.g. the vs. a) or to guide the integration of new information into the previous discourse (e.g. too). Syntactic and prosodic structures (e.g. left dislocation or a particular pitch accent) also provide crucial, conventionalized information. When one starts using distributional models, this distinction between content words and function words cannot be obviated in the way that it has been relatively easy to obviate in formal semantic theories. This fact raises a number of challenges. If the conventional contributions of function words cannot be represented in distributional models in the same way as those of content words, how should they be represented? Some expressions, such as prepositions, manifest properties both of content words and function words; how do we analyze these? These are questions that the main natural language processing applications using distributional models have been able to ignore so far, but continuing to ignore them will probably impose an upper limit on the quality of NLP applications. We might also aspire to having computational models that help us understand human language processing. For example, it would be interesting to see what our models predict for patterns of semantic change, particularly the well-attested phenomenon of semantic bleaching (the loss of descriptive content associated with an expression over time, often substituted by a strictly referential function). Without an analysis of function words, cognitively realistic language models are not possible.

The analysis of meaning and psychological reality: A model of meaning that is cognitively realistic should be compatible with what we know about how language is acquired, how it is processed in real time, how it connects to the rest of our cognitive systems, what happens in pathological situations, how language changes over time, and how we come to associate new or revised concepts with bits of language. Here are just a few disconnected thoughts about language and cognition that have come to my mind as I have worked with distributional semantic models: 1) One appealing thing about distributional semantic models is that they might allow us to avoid making some difficult decisions about the linguistic meaning/world knowledge boundary. 2) Working with distributional models naturally leads one to think of language as decompositional rather than compositional. This change in perspective has all kinds of interesting implications. 3) Distributional models rely on quantities of data that do not correspond to realistic assumptions about exposure to language during development (a point made in a recent paper by Copestake and Herbelot). If these were to map onto cognitively plausible models, clearly more than just raw statistics would have to be influencing their functioning. But what are these other influences, and how do they work? 4) If I have been critical of formal semantics for its almost exclusive emphasis on referential aspects of meaning, I have developed an entirely new appreciation for these – particularly the special

informativity of the association of words with visual or auditory stimuli – when finding them absent in distributional models.

It should be clear that there'll be no shortage of things to talk about during the week

5.9 Incrementality in Compositional Distributional Semantics

Alessandra Zarcone (Universität Stuttgart, DE), Sebastian Padó (Universität Stuttgart, DE)

License © Creative Commons BY 3.0 Unported license
© Alessandra Zarcone, Sebastian Padó

Overall Interest. Our interest is at the crossroads of computational linguistics and psycholinguistics. We are interested in compositional distributional semantic models (CDSMs) that can both contribute towards NLP as well as account for (aspects of) human sentence comprehension. The following ideas come from a project proposal currently in preparation.

Focus and Desiderata. We feel that a promising direction for CDSMs is provided by tensor-based models [3, 6, 2, 1]: each word is associated with one or more types describing its semantic arity and determining the shape of its distributional semantic representation. For example, nouns can be mapped onto vectors, adjectives onto matrices, and verbs and other function words on higher-order tensors. This enables the formulation of syntax-semantics interfaces that look similar to traditional ones but operate on distributional representations, with the potential to link the benefits of distributional representations with the power of compositionality.

Current models though have some limitations: (a) they are constituency-based rather than based on dependencies (dependency grammar is well-established for many languages, in particular with free word order); (b) they are not incremental, that is, they do not construct semantics in a left-to-right manner (whereas human language processing is to a large degree incremental); (c) they do not incorporate a notion of plausibility for (partial) analyses based on expectations at the level of individual composition operations.

We aim at developing a tensor-based CDSM overcoming such limitations. The steps that we foresee are as follows:

1. **A dependency-based distributional syntax-semantics interface.** This step does not yet take incrementality into account. This simplification allows us to binarize the dependency trees of a large German dependency-parsed corpus into composition trees that express the order of semantic composition (see Figure 1); then we will infer the algebraic type(s) of each lemma (nouns as well as sentences are represented as vectors in \mathbb{R}^n , while other parts of speech will generally be assigned higher-order types); we will finally learn a large lexicon that associates lemmas with distributional representations of appropriate algebraic types, via multi-step regression learning [5]. The free choice of binarization schemes allows us to choose one that leads to well-behaved types both in terms of lexical ambiguity and type complexity.
2. **CDSM-based semantic plausibility scores.** Previous definitions of semantic plausibility for predicate argument combinations [4, 7] were limited to predicate-argument combination. They were based on vector similarity, comparing the expectations about arguments against actual arguments. We assume that these approaches can be generalized to our tensor-based CDSMs, with comes with the potential of generalizing semantic plausibility to a wider range of linguistic constructions. Our central assumption is that sentence plausibility decomposes along the edges of the sentence's composition tree.

3. **Incrementality.** The next step is to adapt the first two models to an incremental setup. We will use an incremental dependency parser and assign a semantic representation to each prefix of the sentence that receives a connected analysis from the parser. The central challenge is that in contrast to step 1, we cannot freely choose the order of compositions; instead, the composition tree must be left-branching. This will introduce a considerably higher degree of lexical ambiguity that has to be managed. Subsequently, we want to define incremental plausibility scores by adapting our plausibility measures to the incremental nature of the analysis, taking advantage of the definition of the plausibility measure in terms of individual edges.
4. **Evaluation.** Given a sentence, the model will be able to return plausibility scores for upcoming words at each time during processing. The psycholinguistic evaluation of these scores will take place through word-by-word prediction of reading times. We will perform a broad-scale prediction of reading times on a corpus of German sentences, hoping to show that our plausibility model can account for a larger amount of variance than other models. The NLP-oriented evaluation will be applied to a state-of-the-art beam search-based dependency parser to re-rank dependency parsing hypotheses, both at the level of complete sentences and during parsing.

References

- 1 Baroni, M., Bernardi, R., Zamparelli, R., 2013. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies* (To appear).
- 2 Baroni, M., Zamparelli, R., 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In: *Proceedings of EMNLP*. Cambridge, MA, pp. 1183–1193.
- 3 Clark, S., Pulman, S., 2007. Combining symbolic and distributional models of meaning. In: *Proceedings of the AAAI Spring Symposium on Quantum Interaction*. Stanford, CA, pp. 52–55.
- 4 Erk, K., Padó, S., Padó, U., 2010. A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics* 36 (4), 723–763.
- 5 Grefenstette, E., Dinu, G., Zhang, Y., Sadrzadeh, M., Baroni, M., 2013. Multi-step regression learning for compositional distributional semantics. In: *Proceedings of IWCS*. Potsdam, Germany, pp. 131–142.
- 6 Grefenstette, E., Sadrzadeh, M., Clark, S., Coecke, B., Pulman, S., 2011. Concrete sentence spaces for compositional distributional models of meaning. *Proceedings of IWCS*, 126–134.
- 7 Lenci, A., 2011. Composing and updating verb argument expectations: A distributional semantic model. In: *Proceedings of CMCL*. Portland, Oregon, pp. 58–66.

5.10 Acquiring conceptual knowledge for semantic interpretation

Massimo Poesio (University of Essex, GB)

License  Creative Commons BY 3.0 Unported license
© Massimo Poesio

Summary of my research in the area

Initial Motivations. My first research area was knowledge representation, but I soon stopped working on that to focus on (computational) semantics and pragmatics. I started to look at theories of conceptual knowledge acquisition after a few years working on the use of lexical semantics for anaphora resolution (in particular to interpret bridging references)

and being dissatisfied with the results obtained using WordNet or other lexical resources. Developing theories for acquiring such knowledge automatically seemed both scientifically more interesting and something that was needed to overcome the commonsense knowledge bottleneck found in other areas of AI as well. Over the years however I started getting interested in the acquisition of commonsense knowledge per se and in commonsense knowledge more in general (in particular in cognitive evidence about the way commonsense knowledge is organized).

Acquiring lexical knowledge for resolving bridging references. Our initial efforts were motivated by the work on bridging references carried out with Renata Vieira [18, 17, 24], that had showed that WordNet offered only limited support for this type of interpretation process. Our intuition was that semantic space models like HAL [12] would be quite good at capturing bridging references based on synonymy; our results confirmed this (Poesio et al., 1998). For other types of bridging references we started looking at the unsupervised methods for extracting semantic relations proposed by Hearst [11]. Our work on meronymy indicated that a reasonable precision and recall could be achieved provided that (a) very large corpora were used (Web size), and (b) semantic space models were combined with salience information [20].

Acquisition informed by research on lexical semantics and knowledge representation. As a result of the work on resolving associative references, we started working on theories of commonsense acquisition that incorporated insights from work on lexical semantics (in particular the work by Pustejovsky [23] and formal ontology (in particular the work by Guarino and his lab, [10]). In collaboration with my PhD students Abdulrahman Almuhareb and Eduard Barbu, and then with Marco Baroni, we developed acquisition models that built conceptual relations based on semantic relations extracted from text. With Abdulrahman, we used first unsupervised methods to extract from text attributes, and then supervised methods to build vectors based on qualia theory and Guarino's theory of attributes ([2, 4, 3, 15]). This model also attempted to discriminate between wordsenses ([5]). (A summary of this research can be found in ([16]); a more extensive description in ([1]). With Eduard Barbu, we developed improved models to extract semantic relation-based conceptual descriptions ([21]) and then started using Wikipedia as a corpus ([8]). Finally with Marco Baroni we studied methods using semi-supervised techniques for relation extraction ([9]).

Combining brain evidence with corpus evidence. In recent years, the focus of our research in the area of commonsense knowledge has shifted to using machine learning techniques to study the representation of conceptual knowledge in the brain ([14, 6]) and then using distributional models to predict the activation patterns of concepts ([13, 7]).

Where we stand

At least from a scientific point of view, the only solution to the commonsense bottleneck is to develop models for the acquisition of commonsense knowledge. But the fact remains that although work on using semantic space models for anaphora resolution has continued, the results are still unsatisfactory ([22]). In fact, I would make a more general claim: that so far distributional models have proved successful at tasks that only require collocational or lexical knowledge (checking text coherence, identifying synonymy, etc) but haven't yet been successfully employed in semantic tasks that do require commonsense knowledge. To me the question of why this is the case ought to be one of the central issues for the workshop.

References

- 1 Almuhareb, A. (2006). Attributes in Lexical Acquisition. Ph.D. thesis, University of Essex, Department of Computer Science.
- 2 Almuhareb, A. and Poesio, M. (2004). Attribute- and value-based clustering of concepts. In Proc. of EMNLP , pages 158–165, Barcelona.
- 3 Almuhareb, A. and Poesio, M. (2005a). Concept learning and categorization from the web. In Proc. of Annual Meeting of Cognitive Science Society, pages 103–108, Stresa (Italy).
- 4 Almuhareb, A. and Poesio, M. (2005b). Finding attributes in the web using a parser. In Proc. of the Corpus Linguistics Conference, Birmingham.
- 5 Almuhareb, A. and Poesio, M. (2006). MSDA: A word sense discrimination algorithm. In Proc. of ECAI , Riva del Garda.
- 6 Anderson, A., Murphy, B., and Poesio, M. (2013a). Discriminating taxonomic categories and domains in mental simulations of concepts of varying concreteness. *Journal of Cognitive Neuroscience*.
- 7 Anderson, A., Bruni, E., Bordignon, U., , Poesio, M., and Baroni, M. (2013b). Of words, eyes and brains: Correlating image-based distributional semantic models with neural representations of concepts. In Proc. of EMNLP .
- 8 Barbu, E. and Poesio, M. (2009). Unsupervised knowledge extraction of taxonomies of concepts from wikipedia. In Proc. RANLP , pages 28–32, Borovets.
- 9 Baroni, M., Murphy, B., Barbu, E., and Poesio, M. (2010). Strudel: A distributional semantic model based on property and types. *Cognitive Science*, 34(2), 222–254.
- 10 Guarino, N. (1992). Concepts, attributes and arbitrary relations. *Data and Knowledge Engineering*, 8, 249–261.
- 11 Hearst, M. A. (1998). Automated discovery of wordnet relations. In C. Fellbaum, editor, *WordNet: An Electronic Lexical Database*. MIT Press.
- 12 Lund, K., Burgess, C., and Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In Proc. of the 17th Annual Conference of the Cognitive Science Society, pages 660–665.
- 13 Murphy, B., Baroni, M., and Poesio, M. (2009). Eeg responds to conceptual stimuli and corpus semantics. In Proc. of EMNLP , pages 619–627, Singapore.
- 14 Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., and Lakany, H. (2011). Eeg decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, 117(1), 12–22.
- 15 Poesio, M. and Almuhareb, A. (2005). Identifying concept attributes using a classifier. In T. Baldwin and A. Villavicencio, editors, *Proc. of the ACL Workshop on Deep Lexical Semantics*, pages 18–27, Ann Arbor, Michigan.
- 16 Poesio, M. and Almuhareb, A. (2008). Extracting concept descriptions from the web: The importance of attributes and values. In P. Buitelaar and P. Cimiano, editors, *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pages 28–44. IOS Press, The Netherlands.
- 17 Poesio, M. and Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2), 183–216. Also available as Research Paper CCS-RP-71, Centre for Cognitive Science, University of Edinburgh.
- 18 Poesio, M., Vieira, R., and Teufel, S. (1997). Resolving bridging references in unrestricted text. In R. Mitkov, editor, *Proc. of the ACL Workshop on Operational Factors in Robust Anaphora Resolution*, pages 1–6, Madrid. Also available as HCRC Research Paper HCRC/RP-87, University of Edinburgh.
- 19 Poesio, M., Schulte im Walde, S., and Brew, C. (1998). Lexical clustering and definite description interpretation. In Proc. of the AAAI Spring Symposium on Learning for Discourse, pages 82–89, Stanford, CA. AAAI.

- 20 Poesio, M., Mehta, R., Maroudas, A., and Hitzeman, J. (2004). Learning to solve bridging references. In Proc. of ACL, pages 143–150, Barcelona.
- 21 Poesio, M., Barbu, E., Giuliano, C., and Romano, L. (2008). Supervised relation extraction for ontology learning from text based on a cognitively plausible model of relations. In Proc. of ECAI Workshop on Ontology Learning and Population, pages 1–5, Patras.
- 22 Ponzetto, S., Versley, Y., and Poesio, M. (2014). Using lexical and commonsense knowledge for anaphora resolution. In M. Poesio, R. Stuckardt, and Y. Versley, editors, *Anaphora Resolution: Algorithms, Resources and Applications*, chapter 2. Springer.
- 23 Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.
- 24 Vieira, R. and Poesio, M. (2000). An empirically-based system for processing definite descriptions. *Computational Linguistics*, 26(4), 539–593.

5.11 Research overview

Tim Van de Cruys (Paul Sabatier University – Toulouse, FR)

License  Creative Commons BY 3.0 Unported license
© Tim Van de Cruys

My research has explored different algorithms for the modeling of semantic phenomena within the framework of distributional semantics, with a focus on factorization algorithms and tensor algebra. Below is an overview of the research that is most connected to the seminar themes.

Word meaning in context

An important part of my research focuses on factorization, and its application to language. The use of large text collections brings about a large number of contexts in which a word occurs. By using a factorization algorithm, the abundance of individual contexts can be automatically reduced to a limited number of significant dimensions. Characteristic for these dimensions is that they contain latent semantics: the value of a word on a particular dimension indicates the score of the word for a particular semantic field. This is particularly useful for dealing with polysemous words. By determining the latent semantic fingerprint for a particular context, it is possible to weight the word vector accordingly, thus computing the specific meaning of a word in a particular context [3].

Modeling compositionality

Most research in distributional semantics uses matrices as its main mathematical tool, which is useful for the modeling of individual words. If, on the other hand, one wants to model interactions between several words, multi-way co-occurrences need to be taken into account. Multi-way co-occurrences need to be represented within a tensor framework, which is the generalization of a matrix for more than two modes. Tensors may contain any number of n modes. This allows for the treatment of more complex syntactic constructions, such as the combination of a verb and its different complements, or the different modifiers that a verb appears with. Tensors can equally be combined with factorization algorithms, and they can subsequently be used for the modeling of compositional phenomena [4]. The key idea is that compositionality is modeled as a multi-way interaction between latent factors, which are automatically constructed from corpus data. The model can be readily applied to transitive phrases, for which it gives good results.

Position statement

The opposition that exists between distributional approaches and formal approaches to semantics is very much related to the opposition between connectionist and symbolic models within the field of cognitive science; in a way, they provide two different perspectives on the same data. While formal semantics provides a framework for the explicit, symbolic modeling of semantic phenomena, distributional semantics provides a way to deal with those phenomena in a more implicit way, based on simple co-occurrence data. Formal semantics is typically characterized as very successful with respect to the semantic modeling of functional elements and quantification (elements typically not tackled by the distributional approach), while distributional semantics is lauded for its ability to cope with lexical semantics (which is less extensively developed within the formal semantic framework). Yet, nothing seems to prevent the formal or the distributional approach to model the kind of semantic phenomena that are typically more successfully modeled within the other approach. Distributional models are able to get at the generalizations that are typically handled within a formal semantic framework, while nothing prevents the formal semantic approach from explicitly modeling lexical semantics (though the manual modeling of the lexical semantics of individual content words would quickly become a tedious and prohibitively expensive tasks).

Does this mean that one approach should take precedence over the other? Most likely, the best results are obtained by taking a hybrid approach. The ability of the distributional approach to induce generalizations automatically from corpus data is a huge advantage over the manual approach of formal semantics, while the latter provides machinery for inference and entailment which are still problematic within a distributional framework. What exactly should be the role of each framework is a very interesting topic of discussion, that will probably be amply touched upon during the seminar.

References

- 1 Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In ICLR 2013, 2013.
- 2 Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. Semantic compositionality through recursive matrix-vector spaces. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 1201–1211, Jeju Island, Korea, July 2012. Association for Computational Linguistics.
- 3 Tim Van de Cruys, Thierry Poibeau, and Anna Korhonen. Latent vector weighting for word meaning in context. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 1012–1022, Edinburgh, Scotland, UK, July 2011. Association for Computational Linguistics.
- 4 Tim Van de Cruys, Thierry Poibeau, and Anna Korhonen. A tensor-based factorization model of semantic compositionality. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1142–1151, Atlanta, Georgia, June 2013. Association for Computational Linguistics.

5.12 Semantics, Communication, and Probability

Jan van Eijck (CWI – Amsterdam, NL)

License  Creative Commons BY 3.0 Unported license
© Jan van Eijck

Logic, Linguistics, and Intelligent Interaction. In logic the distinctions between language, interpretation and communication are quite clear, in natural language understanding less so. But maybe natural language semantics has something to learn from new directions in logic. A first lesson was taught by Richard Montague long ago, but there are some new things to learn now.

Logic, narrowly conceived, is the design and use of formal languages for thought, the study of their strengths and limitations (the trade-off between expressive power and complexity), and the use of these tools in clarifying what goes on in the mind of a mathematician, or in the memory of a computer carrying out a program. Montague’s lesson for NL understanding was that NL can be studied with the methods from logic.

Broadly conceived, logic is the study of intelligent interaction, rational adjustment on the basis of evidence, transformation of our conceptualisations of the world on the basis of received information. See [1] for an overview, and for a logic textbook emphasizing this broader perspective.

Intelligent interaction is also a central topic in natural language understanding, for intelligent interaction is what natural language is for. A desire to explain why human beings are so good at communication using language is one of the reasons for being interested in linguistics.

Formal Models of Communication. In dynamic epistemic logic (see [3] or [2]), a state of affairs is a multi-agent Kripke model, and acts of communication are operations on states of affairs. The Kripke model represents what the agents know (or believe). If an agent a is uncertain about the truth of p , this is represented by an inability of a to distinguish p -worlds from non p -worlds. The act of communication represents how this knowledge (or this belief) gets changed by information exchange. A paradigm example is public announcement. A public announcement of a true fact p has the following effect on a Kripke model. All non p -worlds get removed from the model, and the accessibility relations representing the knowledge or belief of the agents get restricted to the new class of worlds. The result is that p becomes common knowledge among all agents. But many other kinds of communication can be modelled: messages to specific individuals, messages to all agents that happen to pay attention, and so on.

Knowledge, Belief, and Probability. In epistemic/doxastic logic (the logic of knowledge and belief), there is also a new trend, where knowledge and belief are linked to probability theory. Theories of subjective probability [6] agree well with Kripke model representations of knowledge and belief. To turn a Kripke model into a probabilistic model, all one has to do is to add, for each agent, a probability distribution over the set of all worlds to the model [5]. Knowledge of a cannot be linked to certainty: assigning probability 1 to a statement. Belief can be linked to assigning probability $> 1/2$ to a statement. This way, it is possible to explain certain properties of belief that are hard to cope with without bringing in probabilities.

Connection with Natural Language Semantics. Probabilistic semantics for natural language would link language (content words) to the world in a loose way (looser than the traditional truth-functional way), in the perspective of an agent (here is where subjective probabilities of the “knowing subject” come in). Example: vague or uncertain attribution.

“Bonfire is black”. In a probabilistic Kripke model M , in a world w for an agent a , this gets a probability $P_{a,w}$. If the probability is 1, this means that a knows that Bonfire is black, and it follows that it is true that Bonfire is black. In a case where the statement is judged as less than certain by a , we can say that a believes that Bonfire is black. Now it does not follow that it is true that Bonfire is black. Program Work out a probabilistic multi-agent semantics for natural language along these lines. See [4] for a first sketch. Connect up with work on distributional semantics.

References

- 1 Benthem, J. v., van Ditmarsch, H., van Eijck, J., and Jaspars, J. Logic in Action. Internet, 2013. Electronic book, available from <http://www.logicinaction.org>.
- 2 Benthem, J. v., van Eijck, J., and Kooi, B. Logics of communication and change. *Information and Computation* 204, 11 (2006), 1620–1662.
- 3 Ditmarsch, H. v., van der Hoek, W., and Kooi, B. *Dynamic Epistemic Logic*, vol. 337 of Synthese Library. Springer, 2006.
- 4 Eijck, J. v., and Lappin, S. Probabilistic semantics for natural language. To appear in the LIRA Yearbook 2013.
- 5 Eijck, J. v., and Schwarzentruher, F. Epistemic probability logic simplified. Submitted to AAMAS 2014, Paris.
- 6 Jeffrey, R. *Subjective Probability – The Real Thing*. Cambridge University Press, 2004.

5.13 Position statement

Dominic Widdows (Microsoft Bing – Bellevue, US)

License  Creative Commons BY 3.0 Unported license
© Dominic Widdows

My interest in compositional semantics and distributional models began when working on the Stanford Infomap project in the early 2000s, and has continued ever since. After stints at MAYA Design, Google, and Bing, I’m now Director of Language Engineering at Serendipity, a startup with the goal of consumerizing analytics, in which the need for good models for compositional semantics is more pressing than ever!

A differential geometer by training, I had the good fortune to work in an area where tensor products, exterior algebra, linear spans and orthogonal complements are widely used, long before realising that these mathematical models and operations could also be applied to natural language. My early adventures in this space included the use of orthogonal complements for negation and linear sum for disjunction in distributional models built using Latent Semantic Analysis.

We released the software implementation of this work as part of the Infomap NLP package, which after a few years was superseded by the SemanticVectors package, which is freely available at <http://code.google.com/p/semanticvectors>. This in turn led to many collaborations, most notably with Trevor Cohen of the University of Texas Health Sciences Center in Houston. Together, we’ve used the package for literature based discovery, drug repurposing, and most recently, orthographic encoding.

The work on drug repurposing and orthographic encoding highlights two important points for the seminar:

- Distributional models can successful for purposes way beyond the pioneering cases in information retrieval and text classification. In the application to drug repurposing, for

example, they are used much more like a fast, robust, approximate theorem prover.

- These models depend on composition operators that are more varied than the simple vector sum. Their success is partly due to the ready availability of established algebraic methods including orthogonal projection, tensor algebra and matrix multiplication, circular convolution, and permutation.

When applied to vectors with complex or binary numbers as coordinates, these operations, their implementations, and experimental results sometimes differ markedly from those obtained with real numbers as coordinates. This points out a sometimes surprising gap in information retrieval and indeed machine learning: in these rapidly developing empirical fields, we tend to tacitly assume that real numbers are the canonical ground field. This is in marked contrast to physics, where complex numbers are ubiquitous, and logic, where binary numbers are the established starting point. One ongoing personal goal of mine is to encourage theoretical and practical researchers in computational semantics to experiment much more with complex and binary vectors as well as real vectors, in the hope that such investigations may prove as fruitful for information retrieval as they have been for physics and logic.

5.14 Norms and Exploitations in Text Meaning and Word Use

Patrick Hanks (University of Wolverhampton, GB)

License  Creative Commons BY 3.0 Unported license
© Patrick Hanks

It is a truism that meaning depends on context. Corpus evidence shows that normal contexts can be summarized and quantified, revealing the platforms of phraseological norms on the basis of which we communicate with one another (i.e. on the basis of which future meanings may be created). A contrasting but equally important discovery is the fact that the potential for creative exploitations of normal contexts by ordinary language users far exceeds anything that has been dreamed up in speculative linguistic theory. These contrasting aspects of words in use are analysed in [2].

Meanings can be seen as evanescent interpersonal cooperative events that take place between speaker and hearer (or, with displacement in time, between writer and reader). They are created by using and exploiting shared knowledge of conventional patterns of word use. As I said publicly for the first time at a Dagstuhl seminar twenty years ago, words in themselves don't have very much meaning—but they do have meaning potential. Different aspects of this potential are activated when words are put into context and used for some real communicative purpose.

“Many if not most meanings require the presence of more than one word for their normal realization.” – [3]

So we may conclude that human linguistic behaviour is indeed rulegoverned, but there is not just a single monolithic system of rules: instead, language use is governed by two interlinked systems: one set of rules governing normal, idiomatic uses of words and another set of rules governing how we exploit those norms creatively. I call this 'the double helix theory of language in use'. It has a profound effect on the ways in which words are distributed across texts. Thirty years of corpus analysis drives us to the conclusions 1) that human languages are a puzzling mixture of logic and analogy and 2) that the importance of analogy in making meanings has been consistently underrated.

Types of creative exploitation include (among others): using anomalous arguments to make novel meanings ellipsis for verbal economy in discourse metaphors, metonymy, and other figurative uses for stylistic effect and other purposes.

The Pattern Dictionary of English Verbs (PDEV; <http://deb.fi.muni.cz/pdev/>; publicly available work in progress) implements this principle by associating meanings with patterns rather than with words in isolation. In PDEV, a pattern consists of a verb and its valencies (otherwise known as ‘clause roles’ or ‘arguments’). Each argument is populated by an open-ended set of lexical items and phrases, which share, to some extent, a semantic value. This means that different senses of a verb can be distinguished according to the semantic values of its arguments. For example, ‘executing an order’ and ‘executing a plan’ go together; they are distinguished from ‘executing a criminal’. These are two different meanings of the same verb, activated by different collocates, even though, structurally, all three examples have identical syntax.

PDEV’s patterns are analogous to the constructions described in Construction Grammar (e.g. [1]). A difference is that PDEV is corpus-driven. Every English verb (and in due course, every predicator—including predicative adjectives) has been or will be analysed on the basis of corpus evidence. Analogous work is in progress in Spanish and Italian.

Each entry in PDEV has the following components:

- A set of syntagmatically distinct patterns (the phraseological ‘norms’)
- An ‘implicature’ (i.e. the meaning and context) for each pattern
- A set of corpus lines illustrating normal uses of each pattern
- Comparative frequencies of each pattern of use of each verb, showing which patterns are most frequent
- A smaller set of corpus lines illustrating creative exploitations, insofar as these are found in the analysed samples
- A shallow ontology of nouns and noun phrases

The CPA shallow ontology serves as a device for grouping together nouns and noun phrases that distinguish one meaning of a verb from another.

References

- 1 Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- 2 Hanks, Patrick. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press.
- 3 Sinclair, John. 1998. The lexical item. In Edda Weigand (ed.), *Contrastive Lexical Semantics*. John Benjamins.

5.15 Position statement

Anna Rumshisky (University of Massachusetts – Lowell, US)

License  Creative Commons BY 3.0 Unported license
© Anna Rumshisky

Personal background

My predilection for distributional approaches stems from the early structuralist education I received as a student of theoretical linguistics in Russia. The analysis of minimal pairs of all kinds and of paradigmatic substitutions in syntagmatic contextual patterns in general was one of the first methodological tools taught to a linguist. Therefore as a computational linguist later in life I found myself aligned with corpus-driven distributional approaches to lexical

analysis. My dissertation work on computational lexical semantics for corpus pattern analysis developed quantitative methods for creating contextualized ad-hoc conceptual categories that I felt were needed the proper handling of selectional preferences.

Current position

I briefly summarize below my position on a couple of issues of interest.

Issue 1. Beyond intratextual distributional patterns. At the core of distributional semantics is the notion of concordance, or a set of contexts that the word appears in. However, as we all know, there is only so much you can gain by looking at the “company” which the word “keeps”. This is not how language is used by humans, and not how it is learned – the language is learned in context provided by the circumstance in which linguistic expressions are uttered. What we need to do is to generalize the notion concordance to include this referential intuition, modeling the circumstance as a set of referents and pragmatic factors of the utterance, including the accompanying actions, participants, participants’ intents, etc.

This is a hefty task, for how do you represent all of these different aspects of the “reality” in which something is uttered? There has been some work on linking linguistic expression to (1) visual information (computer vision) and (2) agents and actions (robotics). But so far, from the point of view of language, at least, it’s mostly been limited to “toy” systems ([1], quite a few of the papers in the recent workshops, cf. [2, 3, 4]).

With respect to modeling context of the utterance, or at least representing it well enough to be able to both record it and do something useful with such recordings, we are at the stage where the distributional analysis of text was back in the 50s. I don’t have a solution for how to usefully model and represent the varied aspects of context, but I do think that this is the direction we need to go, and that we desperately need to take it beyond the toy system stage.

Issue 2. Representing compositionality. One of the issues for DS is the scope. What can we actually usefully do with distributional semantics? We know we can do word meanings, more or less. But what else? Personally, I don’t think compositionality through vector addition or multiplication captures any real linguistic intuition for how meanings are built in composition. A composite linguistic expression is built by virtue of its elements successively restricting further and further the meaning potentials for each other, until a meaning for the composite expression is fixed. In a successful communication, a full sentence has a single interpretation. This needs to be reflected in distributional representation of composite expressions.

References

- 1 H. Yu and J. M. Siskind. Grounded Language Learning from Video Described with Sentences. ACL 2013.
- 2 C. Matuszek, S. Tellex, D. Fox, L. Zettlemoyer. Proceedings of 2012 AAAI Grounding Language for Physical Systems.
- 3 T. Darrell, R. J. Mooney, K. Saenko, eds. Proceedings of NIPS 2011 Workshop on Integrating Language and Vision.
- 4 OD. Goldwasser, R. Barzilay, D. Roth, eds. Proceedings of NAACL 2012 Workshop on Semantic Interpretation in an Actionable Context.

5.16 Position statement

Annie Zaenen (*Stanford University, US*)

License  Creative Commons BY 3.0 Unported license
© Annie Zaenen

I am mainly interested in the inferences that can be drawn from texts. My research focuses on the linguistic elements that license inferences about veridicity and existence: what allows us to conclude that a speaker/author is committed to the view that an event has taken place or that an entity exists?

I collaborate with Cleo Condoravdi, Lauri Karttunen and Stanley Peters and occasional Stanford students in a small research unit at CSLI (Stanford), Language and Natural Reasoning (LNR). We assume that there are in natural language constructions and lexical items that signal that a speaker/author presents an event as factual (veridicity) or an entity as existent with certainty, with a high degree of plausibility, or alternatively allow us to conclude that they are impossible or implausible. It is, however, not easy to classify constructions/lexical items according to these inferential properties because we do not have direct access to speakers/authors intentions and, whatever the linguistic elements that signal the allowed inferences are, they interact with other elements in the discourse context that influence the de facto inferences that hearers/readers draw. We think, contra to de Marneffe, Potts and Manning, that it is important to distinguish between the linguistic components and the real world knowledge components that go into drawing conclusions: generalizations that are made over both together will always be constrained by the specifics of the situation in which they were calculated.

We are trying to develop a methodology that allows us to observe how naive language users draw the inferences we are interested in and to translate this understanding into possible annotations of linguistic material for these inferential properties.

At this point we concentrate on the properties of adjectives with clausal complements. Some of those, especially factive constructions, turn out to be more problematic than existing linguistic literature leads one to believe. First the conditions on the factive uses of specific constructions have not been described in enough detail in the existing literature. Whereas 'It was(n't) stupid of John to leave early' is factive, 'It is(n't) stupid to leave early' is not. In certain cases differences in tense can be associated with rather dramatic differences in interpretation 'He was lucky to break even.' is factive or implicative (see below) but 'He will be lucky to break even.' does not mean that the speaker/author thinks that it is likely or sure that the protagonist will fare well.

Lucky in the future seems to be an idiom and the explanation of the differences between the past and the present tense for impersonal evaluatives will most likely be linked to a better understanding of generic interpretations of the present tense, but in other cases, assumed factive expressions are interpreted as implicative and it is not so easy to decide how they should be treated: From a sentence such as: "I was not brave to venture out" one is supposed to conclude that the speaker did venture out. When one looks at the use of such sentences in context, however, one sees that they are often used as implicatives: the speaker did not venture out. But it is premature to simply conclude that for some speakers (non-native speakers?) brave is an implicative adjective. While this may be true for some of them, experimental evidence suggests that for many (native) speakers the interpretation depends on the context: a sentence such as 'He was not stupid to save money.' gets a factive interpretation, while 'He was not stupid to waste money.' gets an implicative interpretation.

A priori it is not clear how such differences such be accounted for: are the adjectives

ambiguous or are the implicative readings ‘performance’ errors? Our investigations suggest that for some speakers some of the adjectives (most clearly lucky, fortunate and stupid) are ambiguous or even only implicative and that, if the rest the variation has to be treated as a performance error, it is one that is very systematically influence by discourse coherence.

With respect to DS this raises the question whether the approach can distinguish between two readings that have rather closely related lexical environments, my suspicion is that it can in principle but it might need to take much more information into account than is done now. With respect to making the distinction between inferences that arise from the pressure of discourse coherence and those that are due to real lexical ambiguity I would like to see a more general discussion.

6 Panel Discussions

During the week, each group carried out intense discussions on the assigned topics, highlighting potential synergies between distributional and formal semantics, pointing out short term as well as long term strategies to implement them.

Each group prepared and presented a summary of their discussions and proposal. These reports were then unified and harmonized by the seminar organizers. The main results of group work are reported in the sections below.

6.1 Polysemy

Polysemy is a central problem for distributional semantics because typically vector representations do not distinguish word senses. Yet, distributional semantics is likely to be able to provide an important contribution to understand and model phenomena such as polysemy and vagueness. Here are some major challenges that distributional semantic model need to address in the near future:

- Can distributional models distinguish types / senses?
- Are there regularities in the model representations and processes corresponding to regularities in the meaning shifts?
- Can they distinguish productivity and conventionality? can we make the implicit information encoded in the vectors explicit (for example in terms of features and meaning components)?
- Can distributional models be augmented?
- Can we use distributional models to evaluate the analyses of semantic theory, for example analyses of meaning shifts?
- Can the distributional models go beyond that and act as a discovery?
- How can distributional semantics better model the notion of meaning potentials?

6.2 Inference

Inference is a stronghold of formal semantics. Conversely, distributional semantics is still not able to address satisfactorily even the most simple cases of natural language inferences. Here are some major issues concerning the treatment of inference with distributional semantics:

- It is necessary to bridge the gap between formal and distributional notions of inference
- Interesting possibilities might arise from the integration with probabilistic inferences
- One major issue is to what extent is DS able to “tap into” contextual information in text

- It is necessary to collect empirical data about examples of inferences, eventually leading to the creation of shareable datasets for model evaluation:
 - simple items exemplifying specific examples of inferences
 - annotated corpus-based examples

6.3 Compositionality

It makes sense to hypothesize that semantic representations include both something distributional and something “structural”/symbolic. We do not have a single agreed-upon hypothesis of what these mixed or parallel representations should be like:

- Overall, the hypothesis space for what sorts of constituents should have distributional representations is:
 1. distributional representations for words only (and/or words and morphemes)
 2. distributional representations for phrases or perhaps clauses
 3. have both word-level and phrase-level distributional representations available
- We see no reason not to exploit both syntax-driven and discourse-driven composition.
- “Flat” semantic representations for the symbolic side (e.g. Hobbs, MRS, other flat underspecified representations) are an alternative approach to compositionality that may address some of the issues raised by Hinrich Schütze as they are not dependent on the availability of a complete syntactic structure
- It would be ideal if the resulting system was psychologically plausible.
- It would also be ideal if the resulting system were useful for NLP applications.
- We should also look for data sets and problems that will get distributional semantic researchers and formal researchers to talk to each other and benefit from what each approach does significantly better than the other. Examples where DS looks promising include:
 1. co-compositionality (e.g., ‘white wine’)
 2. metonymy
 3. explaining highly context-dependent paraphrases that are below (or beyond) the sense level (so not explainable by a lexical resource)
 4. that part of anaphora that depends on lexical content (e.g. cases of quasi-synonymy like ‘his recent appearance at the Carnegie Hall’/ ‘the concert’ / the evening’)
- It would be interesting to tease apart the influence of discourse dynamics on how we identify referents from its influence on how we interpret lexical items.

6.4 Negation

The group decided to focus on negations, because this is a central aspect of natural language semantics, and yet there is no analysis for it in distributional semantics to date:

- Distributional semantics has no treatment for negation, when viewed in the classical definition;
- Distinction between decontextualized and conversational negation;
- Perspectives for Distributional semantics to help identify the comparison sets for the negated item;
- This approach can possibly link to cognitively inspired models of thought.

7 Next Steps

The seminar organizers together with the participants proposed various activities to carry on the discussions started in Dagstuhl:

- Organize a follow-up meeting (3 days) in Pisa, Italy in September 2014
- Provide details about existing datasets (according to a common format) containing interesting linguistic phenomena, to be used as test set for distributional and formal semantic models
- Groups provide a specification over new datasets for challenging, not yet addressed semantic phenomena
- Groups define annotation metadata for the dataset.
- Groups Identify burning topics for next meeting:
 - a. what is the right architecture?
 - b. information structure
 - c. Finding a task/problem where different areas need to be integrated

Participants

- Nicholas Asher
Paul Sabatier University –
Toulouse, FR
- Marco Baroni
University of Trento, IT
- Peter A. Cariani
Harvard Medical School –
Newton, US
- Stephen Clark
University of Cambridge, GB
- Ann Copestake
University of Cambridge, GB
- Ido Dagan
Bar-Ilan University, IL
- Katrin Erk
University of Texas – Austin, US
- Stefan Evert
Univ. Erlangen-Nürnberg, DE
- Patrick W. Hanks
Univ. of Wolverhampton, GB
- Graeme Hirst
University of Toronto, CA
- Jerry R. Hobbs
Univ. of Southern California –
Marina del Rey, US
- Hans Kamp
Universität Stuttgart, DE
- Lauri Karttunen
Stanford University, US
- Alessandro Lenci
University of Pisa, IT
- Sebastian Löbner
Heinrich-Heine-Universität
Düsseldorf, DE
- Louise McNally
UPF – Barcelona, ES
- Sebastian Padó
Universität Stuttgart, DE
- Massimo Poesio
University of Essex, GB
- James Pustejovsky
Brandeis Univ. – Waltham, US
- Anna Rumshisky
University of Massachusetts –
Lowell, US
- Hinrich Schütze
LMU München, DE
- Mark Steedman
University of Edinburgh, GB
- Suzanne Stevenson
University of Toronto, CA
- Tim van de Cruys
Paul Sabatier University –
Toulouse, FR
- Jan van Eijck
CWI – Amsterdam, NL
- Dominic Widdows
Microsoft Bing – Bellevue, US
- Annie Zaenen
Stanford University, US
- Alessandra Zarcone
Universität Stuttgart, DE



Synchronous Programming

Edited by

Stephen A. Edwards¹, Alain Girault², and Klaus Schneider³

1 Columbia University, US, sedwards@cs.columbia.edu

2 INRIA, FR, alain.girault@inria.fr

3 TU Kaiserslautern, DE, klaus.schneider@informatik.uni-kl.de

Abstract

Synchronous programming languages are programming languages with an abstract (logical) notion of time: The execution of such programs is divided into discrete reaction steps, and in each of these reactions steps, the program reads new inputs and reacts by computing corresponding outputs of the considered reaction step. The programs are called synchronous because all outputs are computed together in zero time within a step and because parallel components synchronize their reaction steps by the semantics of the languages. For this reason, the synchronous composition is deterministic, which is a great advantage concerning predictability, verification of system design, and embedded code generation. Starting with the definition of the classic synchronous languages Esterel, Lustre and Signal in the late 1980's, the research during the past 20 years was very fruitful and led to new languages, compilation techniques, software and hardware architectures, as well as extensions, transformations, and interfaces to other models of computations, in particular to asynchronous and hybrid systems.

This report is a summary of the Dagstuhl Seminar 13471 “Synchronous Programming”, which took place during November 18-22, 2013, and which was the 20th edition of the yearly workshop of the synchronous programming community. The report contains the abstracts of the presentations given during the seminar in addition to the documents provided by the participants on the web pages of the seminar¹.

Seminar 18.–22. November, 2013 – www.dagstuhl.de/13471

1998 ACM Subject Classification C.4 Performance of Systems, D.1.3 Concurrent Programming, D.2.2 Design Tools and Techniques, D.2.4 Software/Program Verification, D.3.3 Language Constructs and Features, D.4.7 Organization and Design, F.3.1 Specifying and Verifying and Reasoning about Programs, F.3.2 Semantics of Programming Languages

Keywords and phrases Synchronous Languages, Hybrid Systems, Formal Verification, Models of Computation, WCET Analysis, Embedded Systems

Digital Object Identifier 10.4230/DagRep.3.11.117

Edited in cooperation with Manuel Gesell

¹ See <http://www.dagstuhl.de/13471> for more information.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Synchronous Programming, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 117–143

Editors: Stephen A. Edwards, Alain Girault, and Klaus Schneider



DAGSTUHL
REPORTS

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Stephen A. Edwards

Alain Girault

Klaus Schneider

License  Creative Commons BY 3.0 Unported license
© Stephen A. Edwards, Alain Girault, and Klaus Schneider

Model-based Design of Embedded Systems

In general, the *development of embedded systems* is a challenging task: Concerning the hardware platforms, developers have to cope with tight resource constraints, *heterogeneous and application-specific hardware architectures*, virtual prototypes, and many other difficulties during the design phases. Concerning the software side, several concurrent tasks are executed on the available hardware, either with or without the help of special operating systems, sometimes statically or dynamically scheduled to the available *hardware platforms*, and sometimes tightly coupled with the hardware platforms themselves (implementing memory barriers etc). Finally, many non-functional aspects have to be considered as well like the energy consumption, the *reliability*, and most important the *prediction of the worst-case computation times*. As many embedded systems are *real-time systems*, it is not sufficient to perform the right computations; in addition, the results have to be available at the right point of time to achieve the desired functionality. Besides, the direct interaction with other systems that often have a continuous behavior requires to consider *cyber-physical systems*. Since many embedded systems are used in safety-critical applications, incorrect or delayed behaviors are unacceptable, so that *formal verification* is often applied. Since, moreover, the development costs have to be minimized, new design flows that allow the development of safe and flexible embedded systems are of high interest.

For these reasons, *model-based design flows* became popular where one starts with an abstract model of the embedded system. Many languages are discussed for such model-based approaches, but most of them are based on only a few models of computation. A *model of computation* thereby defines *which, when and why an action* of the system takes place taking into account the *timeliness*, the *causality*, and the *concurrency* of the computations. Classic models of computation are *dataflow process networks*, where computations can take place as soon as sufficient input data is available, *synchronous systems*, which are triggered by clocks, *discrete-event based systems*, where each process is sensitive to the occurrence of a set of certain events, and *cyber-physical systems* whose behavior consists of discrete and continuous transitions (the latter are determined by differential equations).

It is not surprising that all models of computation have their advantages and disadvantages. For example, dataflow process networks can be naturally mapped to distributed systems and have a robust form of asynchronous concurrency provided that the nodes implement continuous functions (as required for Kahn networks). Synchronous systems are the perfect choice for implementing deterministic systems with predictable real-time behaviors on platforms having a local control (like clocks in digital hardware circuits). Discrete-event based systems are ideal for efficiently simulating systems, since the events directly trigger the next actions.

Many years of research were necessary to understand the above mentioned models of computation in depth to be able to develop corresponding programming languages, compilers and verification techniques. The synchronous programming community made substantial progress in this area: Today, the synchronous programming languages have precise formal

semantics which are supported by efficient compiler techniques. Moreover, synchronous languages provide high-level descriptions of real-time embedded systems so that all relevant requirements for a model-based design flow are fulfilled. There are also graphical versions of these textual languages, notably Safe State Machines (developed from Argos and SyncCharts), and there are commercial versions like SCADE. The SCADE tool provides a code generator certified against DO 178-B, which makes it particularly attractive for the aircraft sector.

Quoting Benveniste et al.: *Today, synchronous languages have been established as a technology of choice for modeling, specifying, validating, and implementing real-time embedded applications. The paradigm of synchrony has emerged as an engineer-friendly design method based on mathematically sound tools* [Proceedings of the IEEE, January 2003].

Open Problems

Despite the incredible progress made in the past, even the combination of the classic synchronous languages Esterel, Lustre, and Signal is not yet fully understood. All these languages are based on the abstraction of physical time to a logical time, where each logical step of time may consist of finitely many executions of actions that are – at least in the programming model – executed in zero time. Such a logical step of the computation matches naturally with an interaction of a reactive system with its environment. However, looking at the details, one can observe that the semantics differ: for example, Lustre and Signal are not based on a single clock like Esterel, and while Esterel's and Lustre's semantics are operational and can therefore be defined by least fixpoints, Signal is rather declarative and requires a more complicated analysis before code generation.

Since different models of computation have different advantages and disadvantages, their *combination* becomes more and more important. This does also imply the *translation and communication between models of computations*. For example, so-called globally asynchronous, locally synchronous (GALS) systems have been developed, mixing both asynchronous and synchronous computations. For model-based designs starting from synchronous languages, special forms of synchronous systems have been defined in terms of the (weakly) endochronous systems. Intuitively, endochronous systems are synchronous systems that can determine from which input ports the values are expected for the next reaction step (and therefore they can derive the absence of other inputs, and they do not need the explicit knowledge of absence). For this reason, one can integrate endochronous systems in an asynchronous environment without destroying their synchronous behaviors.

Similar techniques are used for generating distributed systems from high-level descriptions (like synchronous programs) which lead, e.g., also to first approaches to multithreaded code generation from synchronous languages, which becomes more important due to the advent of multicore processors in embedded system design. More progress is needed and will likely be available in the near future in combining these different forms of discrete models of computations.

The combination of synchronous, endochronous, or asynchronous discrete systems with continuous behaviors to describe cyber-physical systems is still in its infancies. Of course, there are many languages for modeling, simulating, and even formally verifying these systems, but most of these languages lack of a formal semantics, and essentially none of them lends itself for a model-based design like synchronous languages. The generalization of the concepts of synchronous systems to polychronous systems, and even further to cyber-physical systems will be a challenge for future research.

Results of the Seminar

The major goal of the seminar was therefore to allow researchers and practitioners in the field of models of computation and model-based design to discuss their different approaches. Desired results are new combinations of these techniques to form new language concepts and design flows that are able to choose the best suited language for particular components and that allow engineers the sound integration of synchronous and asynchronous, discrete and continuous, or event- and time-triggered systems. Besides this, still more research is required for further developing compilation techniques for future manycore processors, and even to develop special processors like the PRET architectures to obtain better estimated time bounds for the execution of programs.

The seminar proposed here aims at addressing all of these questions, building on a strong and active community and expanding its scope into relevant related fields, by inviting researchers prominent in model-based design, embedded real-time systems, mixed system modeling, models of computation, and distributed systems. The seminar was held in the tradition of the Synchronous Programming (SYNCHRON) workshops that are used as the yearly meeting place for the community in this exciting field. The SYNCHRON workshops started in 1994 at Schloss Dagstuhl, and we were proud to celebrate the 20th edition of the workshop from November 18–22, 2013 again in Schloss Dagstuhl. Previous editions of the SYNCHRON workshop were organized at the following locations:

- 2012:** Le Croisic, France – <http://synchron2012.inria.fr>
- 2011:** Dammarie-les-Lys – <http://synchron2011.di.ens.fr>
- 2010:** Fréjus, France – <http://www.artist-embedded.org/artist/Synchron-2010,2206.html>
- 2009:** Dagstuhl, Germany – <http://www.dagstuhl.de/09481>
- 2008:** Aussois, France – <http://synchron2008.lri.fr>
- 2007:** Bamberg, Germany
- 2006:** L’Alpe d’Huez, France – <http://www.inrialpes.fr/Synchron06/>
- 2005:** Qwara, Malta – <http://www.cs.um.edu.mt/~synchron05/>
- 2004:** Dagstuhl, Germany – <http://www.dagstuhl.de/04491>
- 2003:** Luminy, France – <http://www-verimag.imag.fr/PEOPLE/Nicolas.Halbwachs/SYNCHRON03/>
- 2002:** La Londe les Maures, France – <http://www-sop.inria.fr/tick/Synchron2002.html>
- 2001:** Dagstuhl, Germany – <http://www.dagstuhl.de/01491>
- 2000:** Saint-Nazaire, France
- 1999:** Hyères, France – <http://www-sop.inria.fr/meije/synchron99/location.htm>
- 1998:** Gandia, Spain
- 1997:** Roscoff, France
- 1996:** Dagstuhl, Germany – <http://www.dagstuhl.de/9650>
- 1995:** Luminy, France
- 1994:** Dagstuhl, Germany – <http://www.dagstuhl.de/9448>

During its 20 years of existence, the workshop has significantly evolved: its scope has grown to expand to many languages and techniques that are not classically synchronous, but have been substantially influenced by the synchronous languages’ attention to timing, mathematical rigor, and parallelism. Also, while many of the most senior synchronous language researchers are still active, many younger researchers have also entered the fray and have taken the field in new directions. We carefully selected the potential persons to be invited in that senior and junior researchers of the different branches mentioned above will participate the seminar.

This year, we had 44 participants where 23 came from France, 10 from Germany, 5 from the USA, 2 from Sweden, 2 from UK, one from Portugal and one even from Australia. The seminar had 33 presentations of about 45 minutes length with very active discussions². The presentations can be clustered in typical research areas around synchronous languages like

- synchronous and asynchronous models of computation
- hybrid systems
- causality and other program analyses
- compilation techniques
- predictable software and hardware architectures

It was a pleasure to see that the synchronous programming community is still very active in these research fields and that even after 20 years of research, there are still more and more interesting and fruitful results to be discovered. The following sections contains short abstracts of the presentations of the seminar, and further documents were provided by many participants on the seminar's webpage.

February 2014, Albert Benveniste, Stephen A. Edwards, Alain Girault, and Klaus Schneider

² See <http://www.dagstuhl.de/schedules/13471.pdf> for the schedule.

2 Table of Contents

Executive Summary

Stephen A. Edwards, Alain Girault, and Klaus Schneider 118

Overview of Talks

Desynchronization of Synchronous Systems

Yu Bai 124

Representing Spatially Moving Entities using Time-Variant Topologies

Fernando Barros 124

From Quasi-Synchrony to LTTA

Guillaume Baudart 125

BPDF: A Statically Analyzable DataFlow Model with Integer and Boolean Parameters

Vagelis Bebelis 125

Towards Discrete Controller Synthesis for the Reactive Adaptation of Autonomic Systems

Nicolas Berthier 126

A slow afternoon chez PARKAS and a very fast fly (a fun talk)

Timothy Bourke 126

Modelyze: Embedding Equation-Based DSLs

David Broman 127

Index theory for Hybrid DAE Systems

Benoit Caillaud 127

Functioning Hardware from Functional Specifications

Stephen A. Edwards 128

Debugging and Compiler Bootstrapping with an Equation-Based Language Compiler

Peter A. Fritzson 128

Interactive Verification of Cyber-physical Systems

Manuel Gesell 129

Behavioral Equivalence of Transducers under a Fixed Protocol

Dan R. Ghica 130

Timing Through Types

Dan R. Ghica 130

Precise Timing Analysis for Direct-Mapped Caches

Alain Girault 131

When the decreasing sequence fails... Improving fixpoint approximation in program analysis

Nicolas Halbwachs 131

Modal Interface Automata

Gerald Luetzgen 132

Safety Issues in MARTE/CCSL Specifications

Frederic Mallet 132

In Search of a Physical Semantics of Boussinot’s Reactive Model <i>Louis Mandel</i>	133
From Synchronous to Timed Programming <i>Eleftherios Matsikoudis</i>	133
Berry-Constructive Programs are Sequentially Constructive, or: Synchronous Programming from a Scheduling Perspective <i>Michael Mendler</i>	135
SCCharts – Sequentially Constructive Charts <i>Christian Motika</i>	136
A Tale of Two Semantics: Clocked Dimensions in a Multidimensional Language <i>John Plaice</i>	137
Integrated WCET estimation of multicore applications <i>Dumitru Potop-Butucaru</i>	137
A Causality Analysis for Hybrid Modelers <i>Marc Pouzet</i>	138
Timing Analysis Enhancement for Synchronous Program <i>Pascal Raymond</i>	138
Towards a Formal Software Design Methodology for Predictable Embedded Multi-processor Applications <i>Ingo Sander</i>	139
AstraKahn: Coordination programming by extension and refinement of the Kahn model <i>Alex Shafarenko</i>	139
The Coroutine Model of Computation <i>Chris Shaver</i>	140
libDGALS: A Library-based Approach to Design Dynamic GALS Systems <i>Wei-Tsun Sun</i>	141
Compiling SCCharts (and other Sequentially Constructive Programs) to Hardware and Software <i>Reinhard von Hanxleden</i>	141
Conclusions	142
Participants	143

3 Overview of Talks

3.1 Desynchronization of Synchronous Systems

Yu Bai (TU Kaiserslautern, DE)

License © Creative Commons BY 3.0 Unported license
© Yu Bai

Main reference Y. Bai, K. Schneider, “Isochronous Networks by Construction Design,” to appear in Proc. of the Design, Automation and Test in Europe Conf. (DATE’14).

In this presentation, the main threads of methodologies in desynchronization of synchronous systems are discussed.

In the introduction, the three methods: latency-insensitive design, elastic circuits and desynchronization of synchronous programs are covered briefly, followed by their pros and cons. Finally a model-based approach is proposed in order to cover different design properties.

The second part of the talk introduced the simulation of synchronous elastic circuits in SystemC as an application of the proposed model-based approach, where synthesis of elastic modules and elastic channels are presented.

The last part discussed the endo / isochronous systems. Related concepts are compared with examples. Finally a general theorem of correct desynchronization is introduced: if the synchronous system $P = P_1 \parallel \dots \parallel P_n$ (the synchronous composition of n processes) is constructive and clock-consistent, and each process P_i is patient, then the process P can be correctly desynchronized to a GALS system.

3.2 Representing Spatially Moving Entities using Time-Variant Topologies

Fernando Barros (University of Coimbra, PT)

License © Creative Commons BY 3.0 Unported license
© Fernando Barros

The representation of spatially moving systems is a complex task since communication is unstructured, making it difficult to assess what are the entities currently communicating. Given that interaction is mainly governed by the physical location of the entities, the communication pattern changes over time requiring a dynamic topology. To solve this problem we use the Heterogeneous Flow Systems Specification (HFSS), a modular modeling formalism designed to represent hybrid systems with time-variant topologies. We exploit the ability to represent dynamic topologies as an alternative to a representation using publish/subscribe (pub/sub) communication. Additionally, we show that HFSS dynamic topologies can support a large variety of representations taking advantage of the characteristics of the application domains, enabling more expressive and more efficient descriptions of moving entities.

3.3 From Quasi-Synchrony to LTTA

Guillaume Baudart (ENS – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Guillaume Baudart

Joint work of Baudart, Guillaume; Bourke, Timothy; Pouzet, Marc

A Quasi-periodic System is one where every process P is periodic with a nominal period and a jitter. The time between two ticks may thus vary between 'small margins' during an execution:

Signal values are sent across a bus to one-place buffers at a receiver, whence they are sampled periodically.

In his 'cooking book', Paul Caspi showed how to build abstractions for implementing discrete systems on top of this architecture. In later work, with Albert Benveniste and others, he proposed communication protocols for preserving the discrete semantics of signal flows.

We present a brief survey of this work. In particular, we explain the simple relations between the periods and jitters of real-time tasks, and overwriting and oversamplings of values between writers and readers (it's all a matter of fence posts). We generalize (slightly) the idea of quasi-synchronous traces. We also clarify one of the communication protocols by modelling it in the hybrid synchronous language Zelus.

3.4 BPDF: A Statically Analyzable DataFlow Model with Integer and Boolean Parameters

Vagelis Bebelis (INRIA Grenoble – Rhône-Alpes, FR)

License © Creative Commons BY 3.0 Unported license
© Vagelis Bebelis

Joint work of Bebelis, Vagelis; Fradet, Pascal; Girault, Alain; Lavigueur, Bruno

Main reference V. Bebelis, P. Fradet, A. Girault, B. Lavigueur, "BPDF: A statically analyzable dataflow model with integer and boolean parameters," in Proc. of the Int'l Conf. on Embedded Software (EMSOFT'13), pp. 1–10, IEEE, 2013.

URL <http://dx.doi.org/10.1109/EMSOFT.2013.6658581>

Dataflow programming models are well-suited to program many-core streaming applications. However, many streaming applications have a dynamic behavior. To capture this behavior, parametric dataflow models have been introduced over the years. Still, such models do not allow the topology of the dataflow graph to change at runtime, a feature that is also required to program modern streaming applications. To overcome these restrictions, we propose a new model of computation, the *Boolean Parametric Data Flow* (BPDF) model which combines integer parameters (to express dynamic rates) and boolean parameters (to express the activation and deactivation of communication channels). High dynamicity is provided by integer parameters which can change at each basic iteration and boolean parameters which can even change within the iteration.

The major challenge with such dynamic models is to guarantee liveness and boundedness. We present static analyses which ensure statically the liveness and the boundedness of BPDF graphs. We also introduce a scheduling methodology to implement our model on highly parallel platforms and demonstrate our approach using a video decoder case study.

3.5 Towards Discrete Controller Synthesis for the Reactive Adaptation of Autonomic Systems

Nicolas Berthier (INRIA Rennes – Bretagne Atlantique, FR)

License  Creative Commons BY 3.0 Unported license
© Nicolas Berthier

About a decade ago, and due to the ever growing complexity of computer systems, a trend appeared putting forward the automation of the difficult tasks of software systems administration. The software assigned to this work is usually called an Autonomic Management System (AMS). It is composed of software components that evaluate the dynamics of the system under management through measurements (e.g., workload, memory usage), take decisions, and act upon it so that it stays in a set of acceptable states. Some components ensure performance and availability of the system, while others manage the redundancy of its hardware constituents to deal with errors. However, the actual design of such software leads to inconsistencies in the taken decisions, and coordination issues.

First, to tackle this problem, we take a global view and underscore the reactive nature of AMSs. This point of view allows us to suggest a new approach for the design of AMS software, based on synchronous programming and discrete controller synthesis techniques (DCS). They provide us with high-level languages for the specification of the system to manage, as well as means for statically dealing with inconsistencies and coordination issues. We illustrate our approach by applying our design to a realistic multi-tier application, and present an evaluation of its practicality by using a prototype implementation.

We also exploit the preceding modeling use case to identify the needs for extending DCS algorithms to handle quantitative properties. Over a second phase, we introduce ReaX, a new tool currently under development allowing the synthesis of controllers for logico-numerical reactive programs.

3.6 A slow afternoon chez PARKAS and a very fast fly (a fun talk)

Timothy Bourke (ENS – Paris, FR)

License  Creative Commons BY 3.0 Unported license
© Timothy Bourke

Joint work of Bourke, Timothy; Pouzet, Marc

We briefly present a problem posed to use by Rafel Cases and Jordi Cortadella during a lunch organised by Gerard Berry. We propose solutions in the Simulink tool³ and our language Zélus⁴.

Imagine two cars. One starts at Barcelona and travels at 50 km/hr toward Girona—a distance of 100 km. The other starts at Girona and travels at 50 km/hr toward Barcelona. Between the two is a fly travelling at 80 km/hr, initially from Barcelona toward Girona, and changing direction instantaneously whenever it meets either car. There are two questions.

1. How many zig-zags does the fly do during the two hours of travel?
2. Where will the fly be when the two cars reach their destinations?

³ <http://www.mathworks.com/products/simulink/>

⁴ <http://zelus.di.ens.fr>

We first modelled this problem in Simulink. The number of zig-zags, to our great surprise and pleasure, was **42!** [1] (Using R2012a with the ODE45 solver and a relative tolerance of 1×10^{-3} .)

We then modelled the problem in Zélus. This gave an answer of **48**. (Using the Sundials CVODE solver and a custom implementation of the Illinois algorithm.)

Obviously neither answer is correct since the system is not well defined at the instant the cars pass each other. The important questions are whether we should, or even can, statically detect and reject such cases? or stop with an error at runtime?

References

- 1 D. Adams. *The Hitchhiker's Guide to the Galaxy*. Pan Books, 1979.

3.7 Modelyze: Embedding Equation-Based DSLs

David Broman (University of California – Berkeley, US)

License © Creative Commons BY 3.0 Unported license
© David Broman

Joint work of Broman, David; Siek, Jeremy

Main reference D. Broman, J. G. Siek, “Modelyze: a Gradually Typed Host Language for Embedding Equation-Based Modeling Languages,” Technical report, EECS Department, University of California, Berkeley, UCB/EECS-2012-173, June, 2012.

URL <http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-173.html>

Cyber-physical systems combine computations, networks, and physical processes. Modeling and analysis of such systems are vital engineering techniques to manage complexity and enable rapid prototyping. In particular, complex cyber-physical systems are heterogeneous, requiring various model of computations. A key challenge is to provide both expressive modeling capabilities and mechanisms for analyzing these heterogeneous systems. This talk explores a solution to this challenge based on domain-specific embedded languages. We introduce a host language, named Modelyze, in which various domain-specific modeling languages may be embedded. The key features of Modelyze are first-class functions, which provide a mechanism to abstract components of a model, and typed symbolic expressions, to represent and manipulate equations and expressions. The type system for symbolic expressions supports model-level static error checking and provides an automatic lifting translation to provide seamless integration between the host language and the equations represented by symbolic expressions. The type system is based on gradual typing, enabling early static checking for model engineers while providing expressiveness for domain experts.

3.8 Index theory for Hybrid DAE Systems

Benoit Caillaud (INRIA Rennes – Bretagne Atlantique, FR)

License © Creative Commons BY 3.0 Unported license
© Benoit Caillaud

Hybrid systems modelers exhibit a number of difficulties related to the mix of continuous and discrete dynamics and sensitivity to the discretization scheme. Modular modeling, where subsystems models can be simply assembled with no rework, calls for using Differential Algebraic Equations (DAE). In turn, DAE are strictly more difficult than ODE. They require sophisticated pre-processing using various notions of index before they can be submitted

to a solver. In this talk we discussed some fundamental issues raised by the modeling and simulation of hybrid systems involving DAEs. We focused on the following questions:

- What is the proper notion of index for a hybrid DAE system?
- What are the primitive statements needed for a DAE hybrid systems modeler?

The differentiation index for DAE explicitly relies on everything being differentiable. Therefore, generalizations to hybrid systems must be done with caution. We proposed relying on non-standard analysis for this. Non-standard analysis formalizes differential equations as discrete step transition systems with infinitesimal time basis. We could thus bring hybrid DAE systems to their non-standard form, where the notion of difference index can be firmly used.

3.9 Functioning Hardware from Functional Specifications

Stephen A. Edwards (Columbia University – New York, US)

License © Creative Commons BY 3.0 Unported license
 © Stephen A. Edwards
URL <http://www.cs.columbia.edu/~sedwards>

For performance at low power, tomorrow’s chips will be mostly application-specific logic only powered when needed. I propose synthesizing it from the functional language Haskell. My approach – rewriting to a simple dialect that enables a syntax-directed translation – enables parallelization and distributed memory systems. Transformations include scheduling arithmetic operations, replacing recursion with iteration, and improving data locality by inlining recursive types. I am developing a compiler based on these principles.

3.10 Debugging and Compiler Bootstrapping with an Equation-Based Language Compiler

Peter A. Fritzson (Linköping University, SE)

License © Creative Commons BY 3.0 Unported license
 © Peter A. Fritzson
Joint work of Fritzson, Peter A.; Pop, Adrian; Sjoelund, Martin; Asghar, Adeel; Casella, Francesco
Main reference A. Pop, M. Sjölund, A. Asghar, P. Fritzson, F. Casella, “Static and Dynamic Debugging of Modelica Models,” in Proc. of the 9th Int’l Modelica Conference (Modelica’12), pp. 443–454, Linköping Electronic Conference Proceedings, Linköping University Electronic Press, 2012.
URL <http://dx.doi.org/10.3384/ecp12076443>

The high abstraction level of equation-based object-oriented languages (EOL) such as Modelica has the drawback that programming and modeling errors are often hard to find. In this paper we present static and dynamic debugging methods for Modelica models and a debugger prototype that addresses several of those problems. The goal is an integrated debugging framework that combines classical debugging techniques with special techniques for equation-based languages partly based on graph visualization and interaction. The static transformational debugging functionality addresses the problem that model compilers are optimized so heavily that it is hard to tell the origin of an equation during runtime. This work proposes and implements a prototype of a method that is efficient with less than one percent overhead, yet manages to keep track of all the transformations/operations that the compiler performs on the model. Modelica models often contain functions and

algorithm sections with algorithmic code. The fraction of algorithmic code is increasing since Modelica, in addition to equation-based modeling, is also used for embedded system control code as well as symbolic model transformations in applications using the MetaModelica language extension. Our earlier work in debuggers for the algorithmic subset of Modelica caused instrumentation-based techniques which are portable but turned out to have too much overhead for large applications. The new debugger is the first Modelica debugger that can operate without run-time information from instrumented code. Instead it communicates with a low-level C-language symbolic debugger to directly extract information from a running executable, set and remove break-points, etc. This is made possible by the new bootstrapped OpenModelica compiler which keeps track of a detailed mapping from the high level Modelica code down to the generated C code compiled to machine code. The debugger is operational, supports both standard Modelica data structures and tree/list data structures, and operates efficiently on large applications such as the OpenModelica compiler with more than 100 000 lines of code. Moreover, an integrated debugging approach is proposed that combines static and dynamic debugging. To our knowledge, this is the first Modelica debugger that supports transformational debugging and algorithmic code debugging. This presentation also reports on the first bootstrapping (i.e., a compiler can compile itself) of a full-scale EOO (Equation-based Object-Oriented) modeling language such as Modelica. The Modelica language has been modeled/implemented in the OpenModelica compiler (OMC) using an extended version of Modelica called MetaModelica. OMC models the MetaModelica language and is now compiling itself with good performance. Benefits include a more extensible maintainable compiler, also making it easier to add functionality such as the above mentioned debugging support.

References

- 1 Martin Sjölund and Peter Fritzson. Debugging Symbolic Transformations in Equation Systems. In Proceedings of the 4th International Workshop on Equation-Based Object-Oriented Modeling Languages and Tools, (EOOLT'2011), Zürich, Switzerland, Sept 5, 2011. Published by Linköping University Electronic Press, http://www.ep.liu.se/ecp_home/index.en.aspx?issue=056, Sept 2011.
- 2 Adrian Pop, Martin Sjölund, Adeel Asghar, Peter Fritzson, Francesco Casella. Static and Dynamic Debugging of Modelica Models. In Proceedings of the 9th International Modelica Conference (Modelica'2012), Munich, Germany, Sept. 3–5, 2012.
- 3 Martin Sjölund, Peter Fritzson, and Adrian Pop. Bootstrapping a Modelica Compiler aiming at Modelica 4. In Proceedings of the 8th International Modelica Conference (Modelica'2011), Dresden, Germany, March. 20–22, 2011.

3.11 Interactive Verification of Cyber-physical Systems

Manuel Gesell (TU Kaiserslautern, DE)

License © Creative Commons BY 3.0 Unported license
© Manuel Gesell

Joint work of Gesell, Manuel; Li, Xian; Schneider, Klaus;

Cyber-physical systems (CPS) are widely used in safety-critical applications, there is a crucial need for modeling, simulation, and verification. Numerous approaches and tools for CPS verification have already been proposed in the past. Most of them concentrate on model-checking of finite abstractions of restricted classes of CPS. Interactive verification is

an alternative approach that does not suffer from the high complexity of decision procedures, which is well-suited for CPS verification.

Recently, the synchronous Quartz language has been extended for modeling cyber-physical systems, and a corresponding interactive theorem prover AIFProver is currently in development. It combines both model checking and theorem proving ideas, and supports compositional verification. The prototypical version has already been proved to be applicable to large discrete systems and a well-known benchmark of cyber-physical systems. Here, we will demonstrate the capability of the interactive verification approach and tool worked out so far, together with the key techniques remain to be solved in the near future.

3.12 Behavioral Equivalence of Transducers under a Fixed Protocol

Dan R. Ghica (University of Birmingham, GB)

License  Creative Commons BY 3.0 Unported license
© Dan R. Ghica

Joint work of Ghica, Dan R.; Fredriksson, Olle; Al-Zobaidi, Zaid

This talk gives an overview of the “Geometry of Synthesis” programme of research, concerning the synthesis of hardware descriptions from specifications written in higher-order, imperative, recursive, concurrent programming languages. In the context of hardware synthesis we present a new⁵ technique for aggressive minimisation of state machines taking into account constrained environments, which we call “coherent optimisation”. The main properties of the technique (soundness and compositionality) are proved formally using the proof assistant Agda.

3.13 Timing Through Types

Dan R. Ghica (University of Birmingham, GB)

License  Creative Commons BY 3.0 Unported license
© Dan R. Ghica

Joint work of Ghica, Dan R.; Smith, Alex

We introduce a new general notion of resource based on Bounded Linear Logic (BLL) which has the algebraic structure of a semiring. For timing we use a semiring of schedules which are multisets of contractive linear affine transformation. In order to prove the coherence of the type system we describe a categorical model framework. We present, in the concrete case of timing, a simple type inference algorithm based on generating systems of constraints solvable by an SMT such as Zr.

⁵ As far as we know.

3.14 Precise Timing Analysis for Direct-Mapped Caches

Alain Girault (INRIA Grenoble – Rhône-Alpes, FR)

License © Creative Commons BY 3.0 Unported license
© Alain Girault

Joint work of Andalam, Sidharta; Sinha, Roopak; Roop, Partha; Girault, Alain; Reineke, Jan

Main reference S. Andalam, R. Sinha, P. S. Roop, A. Girault, J. Reineke, “Precise timing analysis for direct-mapped caches,” in Proc. of the 50th Annual Design Automation Conf. (DAC’13), pp. 148:1–148:10, ACM, 2013; available as pre-print at HAL.

URL <http://dx.doi.org/10.1145/2463209.2488917>

URL <http://hal.archives-ouvertes.fr/hal-00842368>

Safety-critical systems require guarantees on their worst-case execution times. This requires modelling of speculative hardware features such as caches that are tailored to improve the average-case performance, while ignoring the worst case, which complicates the Worst Case Execution Time (WCET) analysis problem. Existing approaches that precisely compute WCET suffer from state-space explosion. In this paper, we present a novel cache analysis technique for direct-mapped instruction caches with the same precision as the most precise techniques, while improving analysis time by up to 240 times. This improvement is achieved by analysing individual control points separately, and carrying out optimisations that are not possible with existing techniques.

3.15 When the decreasing sequence fails... Improving fixpoint approximation in program analysis

Nicolas Halbwachs (VERIMAG – Grenoble, FR)

License © Creative Commons BY 3.0 Unported license
© Nicolas Halbwachs

Joint work of Halbwachs, Nicolas; Henry, Julien

Main reference N. Halbwachs, J. Henry, “When the decreasing sequence fails,” in Proc. of the 19th Int’l Symp. on Static Analysis (SAS’12), LNCS, Vol. 7460, pp. 198–213, Springer, 2012; available as pre-print at HAL.

URL http://dx.doi.org/10.1007/978-3-642-33125-1_15

URL <http://hal.archives-ouvertes.fr/hal-00734340>

The classical method for program analysis by abstract interpretation consists in computing a increasing sequence with widening, which converges towards a correct solution, then computing a decreasing sequence of correct solutions without widening. It is generally admitted that, when the decreasing sequence reaches a fixpoint, it cannot be improved further. As a consequence, all efforts for improving the precision of an analysis have been devoted to improving the limit of the increasing sequence. In this paper, we propose a method to improve a fixpoint after its computation. The method consists in projecting the solution onto well-chosen components and to start again increasing and decreasing sequences from the result of the projection.

3.16 Modal Interface Automata

Gerald Luettgen (Universität Bamberg, DE)

License © Creative Commons BY 3.0 Unported license
© Gerald Luettgen

Joint work of Luettgen, Gerald; Vogler, Walter

Several modern interface theories for formally modelling and reasoning about component-based, concurrent systems have been built at the crossroads of de Alfaro and Henzinger’s Interface Automata (IA) and Larsen’s Modal Transition Systems (MTS). Two established examples are Nyman et al.s IOMTS and Bauer et al.s MIO, which differ in their view of component compatibility: IOMTS adopts an optimistic view leading to a more permissive parallel composition operator than MIO’s, but has technical shortcomings regarding (non-)monotonicity of refinement and the treatment of internal computation. In addition, both approaches neither consider conjunction on interfaces nor do they allow extending alphabets when refining system components, which are practically desired properties that enable one to specify and design systems incrementally.

This talk presents the novel interface theory Modal Interface Automata (MIA), which addresses the above shortcomings, and discusses MIA’s design decisions, trade-offs and limitations. The reported research is joint work with Walter Vogler of the University of Augsburg, Germany.

3.17 Safety Issues in MARTE/CCSL Specifications

Frederic Mallet (INRIA Sophia Antipolis – Méditerranée, FR)

License © Creative Commons BY 3.0 Unported license
© Frederic Mallet

Joint work of Mallet, Frederic; Millo, Jean-Vivien

The Clock Constraint Specification Language (CCSL) proposes a rich polychronous time model dedicated to the specification of constraints on logical clocks: i.e., sequences of event occurrences. A priori independent clocks are progressively constrained through a set of clock operators that define when an event may occur or not. These operators can be described as labeled transition systems that can potentially have an infinite number of states. A CCSL specification can be scheduled by performing the synchronized product of the transition systems for each operator. Even when some of the composed transition systems are infinite, the number of reachable states in the product may still be finite: the specification is safe. The purpose of this paper is to propose a sufficient condition to detect that the product is actually safe. This is done by abstracting each CCSL constraint (relation and expression) as a marked graph. Detecting that some specific places, called counters, in the resulting marked graph are safe is sufficient to guarantee that the composition is safe.

3.18 In Search of a Physical Semantics of Boussinot's Reactive Model

Louis Mandel (College de France, Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Louis Mandel

Joint work of Aguado, Joaquin; Barros, Fernando; Baudart, Guillaume; Berry, Gérard; Bourke, Timothy; Boussinot, Frédéric; Caillaud, Benoit; de Simone, Robert; Delaval, Gwenaël; Edwards, Stephen; Ghica, Dan; Girault, Alain; Guatto, Adrien; Caillaud, Benoit; Lüttgen, Gerald; Maraninchi, Florence; Mendler, Michael; Pasteur, Cédric; Potop-Butucaru, Dumitru; Pouzet, Marc; Schneider, Klaus; Vuillemin, Jean

Gerard Berry showed earlier in the morning that the semantics of Esterel is given by electricity in circuits. This talk describes the search for a physical semantics of Boussinot's Reactive Model.

The first part of the talk presents the reactive model through the implementation in ReactiveML of the artwork "Carres Noir et Blanc" of Roger Vilder⁶.

The second part presents the five points of the reactive model which guarantee that all programs are causal by construction:

1. add a delay to the reaction to absence,
2. no strong abort,
3. handler of a weak preemption is executed with a delay,
4. add a delay to read the value of a signal,
5. always favour absence of signals.

The last point shows that the reactive model is not a subset of Esterel. For example the following program is not causal in Esterel but is correct in the reactive model: signal *s* in present *s* then emit *s* else ()

Therefore, electrical circuits do not provide a physical implementation of the reactive model. The assumption presented in this talk is that the reactive model can be implemented with circuits running with water instead of electricity.

Finally, a simulator of this kind of circuit implemented in ReactiveML is presented.

3.19 From Synchronous to Timed Programming

Eleftherios Matsikoudis (University of California – Berkeley, US)

License © Creative Commons BY 3.0 Unported license
© Eleftherios Matsikoudis

High-level programming languages have allowed the programmer to ignore the specifics of the underlying execution platform, and focus just on the logic of the intended computational process, effectively decoupling programs from systems. Programs have become models of the systems that execute them. And conditioned on the absence of faults, any two systems executing the same program will have the same behaviour. This is true for sequential programs, and to some extent, for concurrent programs as well. But what about real-time programs?

In a real-time program, the programmer will typically specify the intended timing properties by direct access to the hardware, or use of available drivers specific to the targeted execution platform. The program becomes part of the system, and different programs

⁶ http://www.rogervilder.com/projets/carre_16.html

```

actor Watchdog {
  interface {
    initialize : input channel (& channel unit)
    reset      : output channel unit
  }

  clear : & channel unit

  thread() {
    wait initialize? in [time , ...)
      clear = initialize

    wait *clear? in (time, time + 1.0)
      skip
    else
      reset = {}
  }
}

```

■ **Figure 1** The definition of a trivial watchdog actor in `act`.

are required to specify the same behaviour on different execution platforms. Real-time programming is still today low-level programming.

Our goal is a high-level programming language for timed systems. We use the term “timed” quite liberally to refer to any system that will determinately order its events relative to some physical or logical clock. We are interested in timed systems that are determinate and causal (see [2], [3]).

We present the basic features of a programming language that we call `act`. `act` is an actor-oriented timed programming language. An `act` program starts with the execution of the actor `main`. `main` can *create* other actors to form a dynamically evolving network of conceptually concurrent, memory isolated components that communicate solely through message passing. All actors in a program share a global notion of *logical* time, directly accessible in a program via the keyword `time`. The language allows for polymorphism in the type of `time`. Logical time advances through the use of *temporal* statements, such as `wait`. Non-temporal statements execute in *zero* logical time.

Figure 1 shows the definition of an actor that implements the functionality of a rather trivial watchdog in `act`. The actor consists of

1. a block of channel definitions, making up the interface of the actor,
2. an uninitialized variable definition local to the actor, representing the state of the actor, and
3. the actor’s thread of control, specifying the behaviour of the actor.

Once created, the Watchdog actor will wait until there is an event at the “initialize” channel, including the time instance at which the actor was created. It will then wait until there is an event at the channel whose address the actor was initialized with, and send a reset signal if there is no such event within 1.0 units of time from the time of initialization.

The watchdog example is interesting because it represents a determinate, causal component that does not preserve the prefix relation on discrete-event signals, and thus, cannot be

implemented as a data flow actor.

The theoretical basis for the design of `act`, and specifically, the choice of the temporal statement `wait`, is its completeness over all *synchronous* causal functions on discrete-event signals.

`act` adopts the zero-execution-time hypothesis common to all synchronous programming languages, and many of its constructs are inspired by Esterel. But it allows for time to be dense, and unlike Esterel, treats conditionals as sequential statements in the resolution of causal loops.

By relating logical to physical time, according to the PTIDES paradigm (see [5], [1]), `act` can be used to program real-time systems. The algorithmic approach presented in [4] can be extended to suitably chosen fragments of the language to perform the schedulability analysis necessary for hard real-time applications.

References

- 1 John C. Eidson, Edward A. Lee, Slobodan Matic, Sanjit A. Seshia, and Jia Zou. Distributed real-time software for cyber-physical systems. *Proceedings of the IEEE*, 100(1):45–59, January 2012.
- 2 Eleftherios Matsikoudis and Edward A. Lee. The fixed-point theory of strictly causal functions. Technical Report UCB/EECS-2013-122, EECS Department, University of California, Berkeley, Jun 2013.
- 3 Eleftherios Matsikoudis and Edward A. Lee. On fixed points of strictly causal functions. Víctor Braberman and Laurent Fribourg, editors, *Formal Modeling and Analysis of Timed Systems*, volume 8053 of *Lecture Notes in Computer Science*, pages 183–197. Springer Berlin Heidelberg, 2013.
- 4 Eleftherios Matsikoudis, Christos Stergiou, and Edward A. Lee. On the schedulability of real-time discrete-event systems. In *Embedded Software (EMSOFT), 2013 Proceedings of the International Conference on*, pages 1–15, 2013.
- 5 Yang Zhao, Jie Liu, and Edward A. Lee. A programming model for time-synchronized distributed real-time systems. In *Real Time and Embedded Technology and Applications Symposium, 2007. RTAS '07. 13th IEEE*, pages 259–268, April 2007.

3.20 Berry-Constructive Programs are Sequentially Constructive, or: Synchronous Programming from a Scheduling Perspective

Michael Mendler (Universität Bamberg, DE)

License © Creative Commons BY 3.0 Unported license
© Michael Mendler

Joint work of Aguado, Joaquín; von Hanxleden, Reinhard; Fuhrmann, Insa
Main reference J. Aguado, M. Mendler, R. von Hanxleden, I. Fuhrmann, “Grounding Synchronous Deterministic Concurrency in Sequential Programming,” to appear in Proc. of European Symposium on Programming (ESOP’14), Grenoble, April 2014.

We introduce an abstract value domain $I(D)$ and associated fixed point semantics for reasoning about concurrent and sequential variable valuations within a synchronous cycle-based model of computation. We use this domain for a new behavioural definition of Berry’s causality analysis for Esterel in terms of approximation intervals. This gives a compact and more uniform understanding of causality and generalises to other data-types. We also prove that Esterel’s ternary domain and its semantics is conservatively extended by the recently proposed sequentially constructive (SC) model of computation. This opens the door to a direct mapping of Esterel’s signal mechanism into boolean variables that can be set and reset

arbitrarily within a tick. We illustrate the practical usefulness of this mapping by discussing how signal reincarnation is handled efficiently by this transformation, which is of complexity that is linear in program size, in contrast to earlier techniques that had, at best, potentially quadratic overhead.

3.21 SCCharts – Sequentially Constructive Charts

Christian Motika (Universität Kiel, DE)

License © Creative Commons BY 3.0 Unported license
© Christian Motika

Joint work of von Hanxleden, Reinhard; Duderstadt, Bjoern; Motika, Christian; Smyth, Steven; Mendler, Michael; Aguado, Joaquin; Mercer, Stephen; O'Brien, Owen

Main reference R. von Hanxleden, B. Duderstadt, C. Motika, S. Smyth, M. Mendler, J. Aguado, S. Mercer, O. O'Brien, "SCCharts: Sequentially Constructive Statecharts for Safety-Critical Applications," Technical Report 1311, Christian-Albrechts-Universität zu Kiel, Department of Computer Science, Dec. 2013.

URL http://www.informatik.uni-kiel.de/uploads/tx_publication/tr-1311-bericht.pdf

We present a new visual language, SCCharts, designed for specifying safety-critical reactive systems. SCCharts uses a new statechart notation similar to Harel Statecharts [3] and provides deterministic concurrency based on a synchronous model of computation (MoC), without restrictions common to previous synchronous MoCs like the Esterel constructive semantics [2]. Specifically, we lift earlier limitations on sequential accesses to shared variables, by leveraging the sequentially constructive MoC [4]. Thus SCCharts in short are SyncCharts [1] syntax plus Sequentially Constructive semantics.

The key features of SCCharts are defined by a very small set of elements, the Core SCCharts, consisting of state machines plus fork/join concurrency.

Conversely, Extended SCCharts contain a rich set of advanced features, such as different abort types, signals, history transitions, etc., all of which can be reduced via semantics preserving model-to-model (M2M) transformations into Core SCCharts. Extended SCCharts features are syntactic sugar because they can be expressed by a combination of Core SCCharts features.

On the one hand this eases the compilation and makes it more robust because it reduces its complexity. On the other hand, using Extended SCCharts features, a modeler is able to abstract away complexity of his or her SCCharts model which increases robustness and readability of a model. This approach enables a simple yet efficient compilation strategy and aids verification and certification.

References

- 1 C. Andre. *Semantics of SyncCharts*. Technical Report ISRN I3S/RR-2003-24-FR, I3S Laboratory, Sophia-Antipolis, France, April 2003.
- 2 G. Berry. *The foundations of Esterel*. In G. Plotkin, C. Stirling, and M. Tofte, editors, *Proof, Language, and Interaction: Essays in Honour of Robin Milner*, pages 425-454, Cambridge, MA, USA, 2000.
- 3 D. Harel. *Statecharts: A visual formalism for complex systems*. *Science of Computer Programming*, 8(3):231-274, June 1987.
- 4 R. von Hanxleden, M. Mendler, J. Aguado, B. Duderstadt, I. Fuhrmann, C. Motika, S. Mercer, and O. O'Brien. *Sequentially Constructive Concurrency – A conservative extension of the synchronous model of computation*. In *Proc. Design, Automation and Test in Europe Conference (DATE'13)*, Grenoble, France, March 2013.

3.22 A Tale of Two Semantics: Clocked Dimensions in a Multidimensional Language

John Plaice (The University of New South Wales, AU)

License © Creative Commons BY 3.0 Unported license
© John Plaice

Joint work of Plaice, John; Beck, Jarryd; Mancilla, Blanca; Wadge, William

In 1975, William W. Wadge and Edward A. Ashcroft introduced the language Lucid, in which the value of a variable was a stream. The successors to Lucid took two paths.

The first path, taken by Lustre, was to restrict the language so that a stream could be provided with a timed semantics, where the i -th element of a stream appeared with the i -th tick of the stream's clock, itself a Boolean stream. Today, Lustre is at the core of the Scade software suite, the reference tool for avionics worldwide.

The second path was to generalize the language to include multidimensional streams and higher-order functions. The latest language along this path is TransLucid, a higher-order functional language in which variables define arbitrary-dimensional arrays, where any atomic value may be used as a dimension, and a multidimensional runtime context is used to index the variables.

This talk will show how the two paths are being brought back together, with the introduction of clocked dimensions to TransLucid, thereby allowing for synchronous, reactive programming to take place within the context of a full-fledged higher-order declarative language.

3.23 Integrated WCET estimation of multicore applications

Dumitru Potop-Butucaru (INRIA – Siège, FR)

License © Creative Commons BY 3.0 Unported license
© Dumitru Potop-Butucaru

Joint work of Potop-Butucaru, Dumitru; Puaut, Isabelle

Main reference D. Potop-Butucaru, I. Puaut, "Integrated Worst-Case Execution Time Estimation of Multicore Applications," in Proc. of the 13th Int'l Workshop on Worst-Case Execution Time Analysis, OASICS, Vol. 30, pp. 21–31, Schloss Dagstuhl/Dagstuhl Publishing, 2013.

URL <http://dx.doi.org/10.4230/OASICS.WCET.2013.21>

Worst-case execution time (WCET) analysis has reached a high level of precision in the analysis of sequential programs executing on single-cores. In this paper we extend a state-of-the-art WCET analysis technique to compute *tight* WCETs estimates of parallel applications running on multi-cores. The proposed technique is termed *integrated* because it considers jointly the sequential code regions running on the cores and the communications between them. This allows to capture the hardware effects across code regions assigned to the same core, which significantly improves analysis precision. We demonstrate that our analysis produces tighter execution time bounds than classical techniques which first determine the WCET of sequential code regions and then compute the global response time by integrating communication costs. Comparison is done on two embedded control applications, where the gain is of 21% on average.

3.24 A Causality Analysis for Hybrid Modelers

Marc Pouzet (ENS – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Marc Pouzet

Joint work of Pouzet, Marc; Benveniste, Albert; Bourke, Timothy; Caillaud, Benoit; Pagano, Bruno

Explicit hybrid systems modelers like Simulink/Stateflow allow for programming both discrete- and continuous-time behaviors with complex interactions between them. A key issue in their compilation is the static detection of algebraic or causality loops. Such loops can cause simulations to deadlock, are a source of compilation bugs and prevent the generation of statically scheduled code.

This paper addresses this issue for a hybrid modeling language that combines synchronous Lustre-like data-flow equations with Ordinary Differential Equations (ODEs). We introduce the operator $\text{last}(x)$ for the left-limit of a signal x . This operator is used to break causality loops and permits a uniform treatment of discrete and continuous state variables. The semantics relies on non-standard analysis, defining an execution as a sequence of infinitesimally small steps. A signal is deemed causally correct when it can be computed sequentially and only progresses by infinitesimal steps outside of discrete events. The causality analysis takes the form of a simple type system. In well-typed programs, signals are proved continuous during integration and can be translated into sequential code for integration with off-the-shelf ODE solvers.

The effectiveness of this system is illustrated with several examples written in Zelus, a Lustre-like synchronous language extended with hierarchical automata and ODEs.

3.25 Timing Analysis Enhancement for Synchronous Program

Pascal Raymond (VERIMAG – Grenoble, FR)

License © Creative Commons BY 3.0 Unported license
© Pascal Raymond

Joint work of Raymond, Pascal; Maiza, Claire

In real-time systems, an upper-bound on the execution time is mandatory to guarantee all timing constraints: a bound on the Worst-Case Execution Time (WCET). High-level synchronous approaches are usually used to design hard real-time systems and specifically critical ones. Timing analysis used for WCET estimates are based on the executable binary program. Thus, a large part of semantic information, known at the design level, is lost due to the compilation scheme (typically organized in two stages, from high-level model to C, and then binary code). In this paper, we aim at improving the estimated WCET by taking benefit from high-level information. We integrate an existing verification tool to check the feasibility of the worst-case path. Based on a realistic example, we show that there is a large possible improvement for a reasonable analysis time overhead.

3.26 Towards a Formal Software Design Methodology for Predictable Embedded Multiprocessor Applications

Ingo Sander (KTH Royal Institute of Technology, SE)

License  Creative Commons BY 3.0 Unported license
© Ingo Sander

The presentation addresses the increasing complexity of software design for multiprocessor embedded systems by proposing a design methodology that combines a formal foundation based on the theory of models of computation (MoCs) and the industrial system design language SystemC. In particular a software synthesis flow is presented that starts with an executable system model and yields an implementation on a multiprocessor system-on-chip.

The ForSyDe (Formal System Design) methodology provides the designer with SystemC modeling libraries that lead to executable system models from which abstract analyzable models can be extracted. Using these abstract models, the design space exploration, mapping and synthesis process can make use of the rich set of existing MoC theory by for instance incorporating scheduling and buffer optimization techniques to yield an efficient implementation on a multiprocessor system-on-chip. The presentation will also discuss to what extent performance guarantees can be given provided a predictable architecture is used as target architecture.

3.27 AstraKahn: Coordination programming by extension and refinement of the Kahn model

Alex Shafarenko (University of Hertfordshire, GB)

License  Creative Commons BY 3.0 Unported license
© Alex Shafarenko

This talk introduces some concepts of the coordination language AstraKahn which can be used for programming synchronous and asynchronous systems within the same framework. The talk dwells primarily on the bottom layer of the AstraKhan stack, which is called the Topology and Progress Layer. The concept of pressure-based progress control is explained and the coordination of pressure via state-machine based synchronisers is discussed. This is work in progress. The current definition is available in the form of Arxiv preprint [1].

References

- 1 Alex Shafarenko *AstraKahn: A Coordination Language for Streaming Networks*. arXiv.org arXiv:1306.6029v3 [cs.PL], 2013

3.28 The Coroutine Model of Computation

Chris Shaver (University of California – Berkeley, US)

License © Creative Commons BY 3.0 Unported license
© Chris Shaver

Joint work of Shaver, Chris; Lee, Edward A.

Main reference C. Shaver, E. A. Lee. “The coroutine model of computation,” in R. B. France et al. (eds.) “Model Driven Engineering Languages and Systems”, LNCS, Vol. 7590, pp. 319–334, Springer, 2012.

URL http://dx.doi.org/10.1007/978-3-642-33666-9_21

The Coroutine Model of Computation, defined by Shaver and Lee [5], is a formalism that generalizes other control-oriented models such as state machines, modal models, and imperative programs into a denotational language. Specifically, this denotational language is expressed in terms of the Modular Actor Interfaces of Tripakis et al. [6]. The semantics of this model defines a general interface for Continuation Actors, Actors that in addition to the usual inputs, outputs, and state have control-oriented features: control entry points, control exit locations, and the ability to suspend, terminate, or resume in the context of their containing model. These Continuation Actors can be assembled into a transition system, forming a Coroutine Model.

As opposed to conventional formalisms for state machines, the decision whether or how to transition control, typically codified in a transition guard language, is instead formally part of the interface of each individual Continuation Actor, called its ‘enter’ function. This idea is derived from Andre’s semantics for SyncCharts[1, 2], where he makes a similar association of control decisions with Reactive Cells. Consequently, a simple denotational semantics is given for a Coroutine Model that traverses a sequence of Coroutine Actors, firing each to produce outputs, and deciding how to proceed, whether to suspend, or whether to terminate after executing each Coroutine Actor by using the ‘enter’ function in its interface.

The semantics of the Coroutine Model are additionally extended to accommodate a non-strict form of operation. Given partial information about inputs, in the form of a domain representation, the domain of the power set with inverse inclusion is used to represent partial information about the finite set of possible control decisions for each Continuation Actor. With this definition of partial control decisions at each Continuation Actor, non-strict operation can be defined on the level of the whole Coroutine Model. A simple denotational definition is given for this behavior, and it is proven under this definition that if each contained Continuation Actor defines its outputs and control decisions as monotonic functions of its inputs (in the domain-theoretic sense), the semantics of the whole model will define its outputs and its ultimate control decision as monotonic functions of its inputs.

This monotonicity property is important in the context of hierarchical and heterogeneous models as a form of compositionality. In particular, a collection of such monotonic Coroutine Models can be put together in a mutual constructive fixed-point computation over their connected inputs and outputs, such as that of the Synchronous Reactive model defined by Edwards [3]. This property allows the Coroutine Model of computation to give the semantic quotient of control-oriented synchronous languages over fixed-point computations, providing a theoretical framework for expressing models such as SyncCharts [1] as hierarchical compositions of Coroutine Models and Synchronous Reactive Models, as is done with the KlePto translation of Motika [4].

References

- 1 C. Andre. *SyncCharts: A visual representation of reactive behaviors*. Rapport de recherche tr95-52, Universite de Nice-Sophia Antipolis, France, 1995.

- 2 C. Andre. *Computing synccharts reactions*. Electronic Notes in Theoretical Computer Science, 88(33):3–19, 2004.
- 3 S.A. Edwards. *The Specification and Execution of Heterogeneous Synchronous Reactive Systems*. Technical report, University of California Berkeley, 1997.
- 4 C. Motika. *Semantics and Execution of Domain Specific Models—KlePto and an Execution Framework*. Diploma thesis, Christian-Albrechts-Universität zu Kiel, Department of Computer Science, December 2009.
- 5 C. Shaver, and E. A. Lee. *The coroutine model of computation*. Model Driven Engineering Languages and Systems. Springer Berlin Heidelberg, 2012. 319-334.
- 6 S. Tripakis, C. Stergiou, C. Shaver, and E.A. Lee. *A modular formal semantics for Ptolemy*. Mathematical Structures in Computer Science, 23, pp 834-881. doi:10.1017/S0960129512000278.

3.29 libDGALS: A Library-based Approach to Design Dynamic GALS Systems

Wei-Tsun Sun (INRIA Grenoble – Rhône-Alpes, FR)

License © Creative Commons BY 3.0 Unported license
© Wei-Tsun Sun

We tackle the problem of designing and programming dynamic and reactive systems with four objectives: being based on a formal model of computation, using different types of concurrency, being efficient, and tolerating failures. The challenge lies in the fact that good formal models with very high level of abstraction generally result in non-efficient implementations. We propose a C based library approach following the formal Dynamic Globally Asynchronous Locally Synchronous (DGALS) model of computation. We show how a DGALS system can be dynamically constructed from concurrent behaviors on distributed platforms thanks to the DGALS paradigm. Finally, our experimental results clearly indicate the large execution time and memory footprint gains compared to the current state of the art approaches.

3.30 Compiling SCCharts (and other Sequentially Constructive Programs) to Hardware and Software

Reinhard von Hanxleden (Universität Kiel, DE)

License © Creative Commons BY 3.0 Unported license
© Reinhard von Hanxleden

SCCharts [3] extend SyncCharts [1] with sequential constructiveness (SC) [2] and other features. We developed a compilation chain that first, in a *high-level compilation* phase, performs a sequence of model-to-model transformations at the SCCharts-level [3] such that they can be mapped directly to SC Graphs (SCGs). Then two alternative *low-level compilation* approaches allow mapping to hardware and software; the circuit approach generates a netlist, the priority approach simulates concurrency with interleaved threads.

References

- 1 C. Andre. *Semantics of SyncCharts*. Technical Report ISRN I3S/RR-2003-24-FR, I3S Laboratory, Sophia-Antipolis, France, April 2003.
- 2 R. von Hanxleden, M. Mendler, J. Aguado, B. Duderstadt, I. Fuhrmann, C. Motika, S. Mercer, and O. O'Brien. *Sequentially Constructive Concurrency – A conservative extension of the synchronous model of computation*. In Proc. Design, Automation and Test in Europe Conference (DATE'13), Grenoble, France, March 2013.
- 3 R. von Hanxleden, B. Duderstadt, C. Motika, S. Smyth, M. Mendler, J. Aguado, S. Mercer, and O. O'Brien. *SCCharts: Sequentially Constructive Statecharts for Safety-Critical Applications*. Technical Report 1311, Christian-Albrechts-Universitaet zu Kiel, Department of Computer Science, Dec 2013.

4 Conclusions

The seminar had many high-quality presentations and even more important, many fruitful discussions afterwards until the late evening hours. During the seminar, several research groups discussed their work, and it is not surprising that, as in previous years, some common research has been initiated. For example, Reinhard von Hanxleden and Michael Mendler are currently working on a shared DFG project where they explore new forms of causality that is called sequential causality, which is closer to the traditional sequential programming languages. As another example, Klaus Schneider, Jean-Pierre Talpin and Sandeep Shukla have published several papers about the combination of polychronous and synchronous languages including aspects like clock consistency and causality which cross fertilized both areas. Many other projects benefited from contributions of experts in the area, and therefore we are sure that the Synchronous Programming workshop will celebrate also other anniversaries in the future.

Participants

- Joaquin Aguado
Universität Bamberg, DE
- Mihail Asavae
VERIMAG – Grenoble, FR
- Yu Bai
TU Kaiserslautern, DE
- Fernando Barros
University of Coimbra, PT
- Sanjoy K. Baruah
University of North Carolina –
Chapel Hill, US
- Guillaume Baudart
ENS – Paris, FR
- Vagelis Bebelis
INRIA Grenoble –
Rhône-Alpes, FR
- Gérard Berry
INRIA Sophia Antipolis –
Méditerranée, FR
- Nicolas Berthier
INRIA Rennes – Bretagne
Atlantique, FR
- Timothy Bourke
ENS – Paris, FR
- David Broman
University of California –
Berkeley, US
- Benoit Caillaud
INRIA Rennes – Bretagne
Atlantique, FR
- Albert Cohen
ENS – Paris, FR
- Willem-Paul de Roever
Universität Kiel, DE
- Robert de Simone
INRIA Sophia Antipolis –
Méditerranée, FR
- Gwenaël Delaval
University of Grenoble – LIG, FR
- Stephen A. Edwards
Columbia Univ. – New York, US
- Peter A. Fritzon
Linköping University, SE
- Manuel Gesell
TU Kaiserslautern, DE
- Dan R. Ghica
University of Birmingham, GB
- Alain Girault
INRIA Grenoble –
Rhône-Alpes, FR
- Adrien Guatto
ENS – Paris, FR
- Nicolas Halbwachs
VERIMAG – Grenoble, FR
- Jun Inoue
ENS – Paris, FR
- Xian Li
TU Kaiserslautern, DE
- Gerald Lüttgen
Universität Bamberg, DE
- Antoine Madet
ENS – Paris, FR
- Frédéric Mallet
INRIA Sophia Antipolis –
Méditerranée, FR
- Louis Mandel
College de France, Paris, FR
- Florence Maraninchi
VERIMAG – Grenoble, FR
- Eleftherios Matsikoudis
University of California –
Berkeley, US
- Michael Mendler
Universität Bamberg, DE
- Christian Motika
Universität Kiel, DE
- Valentin Perrelle
IRT SystemX, FR
- John Plaice
The University of New South
Wales, AU
- Dumitru Potop-Butucaru
INRIA – Paris, FR
- Marc Pouzet
ENS – Paris, FR
- Pascal Raymond
VERIMAG – Grenoble, FR
- Ingo Sander
KTH Royal Institute of
Technology, SE
- Klaus Schneider
TU Kaiserslautern, DE
- Alex Shafarenko
University of Hertfordshire, GB
- Chris Shaver
University of California –
Berkeley, US
- Wei-Tsun Sun
INRIA Grenoble –
Rhône-Alpes, FR
- Reinhard von Hanxleden
Universität Kiel, DE



Global Measurement Framework

Edited by

Philip Eardley¹, Marco Mellia², Jörg Ott³, Jürgen Schönwälder⁴,
and Henning Schulzrinne⁵

1 BT – Suffolk, GB, philip.eardley@bt.com

2 Politecnico di Torino, IT, mellia@tlc.polito.it

3 Aalto University, FI, jorg.ott@aalto.fi

4 Jacobs University – Bremen, DE, j.schoenwaelder@jacobs-university.de

5 Columbia University, US, hgs@cs.columbia.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 13472 “Global Measurement Framework”.

Seminar 17.–20. November, 2013 – www.dagstuhl.de/13472

1998 ACM Subject Classification C.2.3 Network Operations, C.4 Performance of Systems

Keywords and phrases Internet measurements, Quality of experience, Network management, Traffic engineering

Digital Object Identifier 10.4230/DagRep.3.11.144

1 Executive Summary

Philip Eardley

Marco Mellia

Jörg Ott

Jürgen Schönwälder

Henning Schulzrinne

License © Creative Commons BY 3.0 Unported license
© Philip Eardley, Marco Mellia, Jörg Ott, Jürgen Schönwälder, and
Henning Schulzrinne

The Internet has a history of unexpected and often unpredictable behaviors due to manifold interactions of thousands of networks, and billions of components and devices and users. The resulting complexity requires measurements to understand how the network is performing, to observe how it is evolving, and to determine where failures or degradations occur. Especially with constantly evolving applications and their interaction paradigms, new phenomena occur and need to be factored into operations and management: one example is the substantial effort going into defining interfaces to assist peer-to-peer applications so that the amount of cross-ISP traffic is reduced. Measurements thus form an integral part of network operator tool sets to keep the net up and running.

But measurements are equally important for the research community to understand network traffic as well as protocol and application dynamics and their evolution. And they assist in quantifying application and (access) network performance and thus provide a tool for end users and regulators to monitor operators and their service level agreements. Tools such as speedtest.net have become widely used for individual measurements and basic ISP rating. Measurement service providers such as SamKnows or RIPE offer networks of probes,



Except where otherwise noted, content of this report is licensed
under a Creative Commons BY 3.0 Unported license

Global Measurement Framework, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 144–153

Editors: Philip Eardley, Marco Mellia, Jörg Ott, Jürgen Schönwälder, and Henning Schulzrinne



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

i.e., separate devices or embedded software on access routers, for continuous background measurements at the end users. These help ISPs and regulators in their work. Standards bodies such as the IETF and the Broadband Forum have established working groups to define a global measurement architecture and common interfaces and to extend the set of metrics describing communication properties.

This Dagstuhl seminar brought together researchers from industry, academia, and regulators to discuss the state of the art in measurements and their exploitation, measurement and analysis techniques, privacy and anonymization, and to contribute to a common understanding in a number of areas, including:

- improving the expressiveness of measurement metrics (and develop appropriate new ones) beyond throughput, loss rate, and RTT so that the actual application-specific user *quality of experience* can be assessed;
- expanding the reach, scale, and diversity of measurements and the corresponding data analysis to obtain a more comprehensive view on the performance of networks and applications;
- structuring the otherwise mostly disconnected measurement activities to allow interfacing between them and/or providing defined access methods to them, for both carrying out measurements and accessing measurement results (offline and in real-time);
- providing ways to better instrument and more broadly utilize measurement infrastructure, inside operators, for end users, and at third parties.

Because the means for taking steps towards achieving the above goals was on learning about and from each other and developing joint perspectives, the seminar chose an extremely interactive organization comprising three elements:

1. Individual presentations were limited to an initial round of introductions (1 slide each) covering a set of questions for the participants to get know each others background and interests.¹
2. Panel discussions (with ample involvement of the “audience”) set the stage for the discussion topics of the day.
3. Extensive group work to dive into a number of topics and also for presenting and discussing the group outcome on the next day.

A side effect of this organization is that there were virtually no individual talks and hence no talk abstracts were collected.

We focused on two complementary aspects of a global measurement framework: 1) *creating* a global measurement framework and 2) *using* such a framework. Both were introduced by panels, with a lot of discussion contributing to these overviews, as described in the respective introduction to the following two sections.

¹ The complete slide set is available on the seminar web page at <http://www.dagstuhl.de/13472/>.

2 Table of Contents**Executive Summary**

<i>Philip Eardley, Marco Mellia, Jörg Ott, Jürgen Schönwälder, and Henning Schulzrinne</i>	144
----------------------------------------------------------------------------------------------------------	-----

Building a Global Measurement Framework

Measurements and metrics	147
Doing it Wrong: Worst Practices	148
Privacy	148
Latency measurements	149
Infrastructure and interfaces	149

Using a Global Measurement Framework

Use Case: Service Provider Trouble Isolation	151
Data Analysis	152

Impressions and Next Steps 152**Participants 153**

3 Building a Global Measurement Framework

Arthur Berger (Akamai), Benoit Claise (Cisco), Sam Crawford (SamKnows), and Daniel Karrenberg (RIPE) introduced aspects and issues of building a (commercially) viable network measurement infrastructure and its constituents, emphasizing their respective angle on the problem. Today, we have many different measurements systems in operation at different scale and reach. Each of these systems exhibits a bias in some form, e.g., whom they are measuring for (ISPs vs. regulators vs. users vs. researchers), whether measurements are active or passive or a combination of both, where the measurement points are located, etc. A truly global measurement framework has to bridge these biases. Different target groups may be interested in different metrics and different granularity of reporting (which may or may not be compatible). Active measurements are good for reference measurements, as the test traffic is defined, but will only offer limited insight as measurement traffic type and timing won't perfectly mimic real users. Passive measurements may be more inclusive; on the other hand, they can only be made on the traffic that exists at that moment and also raise issues concerning user traffic observations and thus privacy concerns. And the number and location of vantage points will be different depending on the questions we are trying to answer. We also covered pitfalls in designing and operating large-scale measurement systems and privacy aspects as cross cutting themes.

One key topic for subsequent discussion, metrics and measurements, was of such broad interest that it was covered in plenary style. In addition, we identified four working groups to follow up on selected sub-topics: 1) Doing it wrong: worst practices, 2) Privacy, 3) Latency measurements, 4) infrastructure and interfaces. We will briefly recap their results below.

3.1 Measurements and metrics

We identified different use cases and, as already mentioned, measurements and metrics may differ depending on which use we are looking at. In general, however, it is important that metrics are clearly defined – how they are measured (tests, vantage points, math, number and intervals of repetitions, time span, etc.) and what their semantics are (what they are useful for and what not) – so that they can be implemented by different parties (in different environments) but yield reproducible and comparable outcomes. To counter implementation errors, reference code may help. To prevent interpretation errors, metrics should be clearly and unambiguously observable. Whenever measurements are carried out and documented in data sets, it is important to record the conditions for the measurements and provide sufficiently detailed documentation in the data sets to allow for later re-use, comparison, etc. without running the risk of misinterpretation.

Concerning standardization, we have to start with the metrics that are understood well enough and are of broad interest; quite a few (simple ones) are already addressed by IETF or ITU specifications. Operational measurement infrastructures (SamKnows, RIPE ATLAS) measure some 15 metrics (as active measurements), focusing on those motivated user and regulatory use cases. Metrics are evolving to become more sophisticated (which creates a tension between standardization and innovation): from measuring simple download speeds and maybe round-trip-times (as is still dominant when judging ISPs) to application-dependent impact (e.g., web page load time, rebuffering events for video streaming) and the more “elusive” quality of experience in general, including availability and robustness. This increasing complexity requires careful and thorough definitions. Many subtleties matter,

which need to be identified: to avoid implementation and measurement errors, to control bias, to include in documentation of results and data sets, etc. The potential for (arbitrary) rise in complexity may call for modesty in what we try to achieve so that metrics and measurement definitions aren't overdone beyond what can be meaningfully documented, interpreted, and compared.

Metrics are used for two closely related yet distinct purposes that would be nice to disentangle: network measurements per se to understand network behavior and benchmarking for performance comparison or rating. The former probably calls for a larger number and more fine-grained metrics, whereas the latter may require fewer metrics and coarser levels.

3.2 Doing it Wrong: Worst Practices

The introduction panel and the general discussion about metrics already captured system design aspects regarding failures. This group work focused on data gathering, processing, documentation, and to some extent interpretation. The first observation emphasized was the importance of metadata (already alluded to above): the environment (e.g., cross-traffic, outages), the context (e.g., applications, usage), and information about the measurement infrastructure (e.g., version and patch numbers). Such metadata are equally important for proper interpretation of results as the measurements themselves.

Ideally the measurement data should always be kept in raw form, not just results after applying an interpretation function, because this function may change over time and one may want to go back and look at old results in the light of new insights. Before measuring, measurement setups (including specific implementations) need to be (repeatedly) calibrated; ideally experiments include such calibration steps. Lack of calibration would lead to uncertainty of the results. Once measurement data is collected, correction factors may have to be applied to raw data and sanity checks (“repair”) should be performed to discard data that are obviously broken. Care needs to be taken to not create a bias during such steps.

Finally, data needs to be interpreted in the context of, e.g., a specific application to understand if the results are “good” or “bad”. A recent trend has been towards developing quality of experience metrics, but this is work in progress and really hard to get right. Particular caution is required when attempting to map observable network characteristics to subjective experience metrics.

3.3 Privacy

Privacy is a tricky topic when it comes to measurements, especially when performing passive measurements (when user traffic patterns or even user data are observed to understand network performance), but also to some extent when carrying out active measurements (e.g., when measurement systems try to avoid colliding with user activity and thus time stamps from measurements yield insights into when users are active). This becomes particularly relevant when data is not (just) used for internal evaluation (e.g., of an ISP or a measurement platform) but when (raw or pre-processed) data is anonymized and shared, e.g., for research. Mistakes, e.g., during the anonymization process cannot be undone (examples from the past can be found in literature).

Defining the problem space requires understanding against whom to protect the users,

businesses, or other entities (e.g., against service providers, applications, governments) and who could serve as the trusted entity offering this protection (e.g., ISPs, governments, privacy service providers, or even the crowd). Measurements and privacy can be at odds with each other (and when data collection is carried out by governmental organization they indeed are, as recent history has shown). Yet, in many cases, people are constantly giving their usage and performance data away to service providers anyway (or they would need to pay for a service). An extreme position would be to argue that, under present circumstances, there is not much of privacy left in the first place. A “privacy as a service” provider could help changing this when interacting with individual services or ISPs, but this would come at a cost for the user.

For data collection using measurement along the lines of this seminar, rules could be defined for maintaining user-related data once collected. The first aspect is minimizing the amount of data collected, anonymizing to a good degree, and then discussing mechanisms how to achieve this. We note that there may be some tension between following these ideas of user privacy protection and keeping raw and encompassing metadata as discussed in the previous section.

3.4 Latency measurements

Measuring latency, while straightforward at the first glance, features numerous sophisticated subtleties when looking at different protocols and applications: latency may be defined in different ways depending on the intent: for example, round-trip time of an ICMP or UDP ping, TCP connection establishment, application layer latency. When concerned with network measurements, we typically try to measure the latency imposed by the network, but other sources of delay exist: in the operating system, in the server or data center, and then in different segments of a network path (home network, access network, etc.).

Latency measurements can be carried out using a number of tools operating at different layers (the simplest ones being *ping* and *traceroute*). They may carry out measurements end-to-end between two hosts or they may receive ISP support (e.g., for timestamping packets when they pass through) so that finer-grained resolution along the measured path becomes possible. Measurements can determine the base RTT (some flavor of calibration), the latency under load (max RTT), and delay variation. Latency-related metrics include RTT and one-way delay variation. Tools for these basic metrics are available.

What is missing includes: being able to identify the source of latency (which requires cooperation of the ISPs), transforming basic latency measurements into semantically richer metrics that reflect the user experience (which is highly application specific and tricky to achieve, as noted above), extensive latency measurements in mobile networks, and support for passive latency measurements at a single point in the network (e.g., when requests and responses or TCP segments can be mapped).

3.5 Infrastructure and interfaces

The group distinguished four different architectures for large-scale measurements, ranging from ubiquitous, but fixed-function devices, to fully programmable custom applications, typically on general-purpose computing equipment. The components of a measurement platform may be owned by the ISP or a dedicated measurement entity or by the subscriber (e.g., third party

modems). Networks to be measured (and whose contributions to measurement results may have to be dissected) include the public networks (ISPs) and private networks (enterprises, universities, etc.)

We differentiate three classes of measurement use cases that differ in scale and purpose:

1. (continuous) large-scale measurements to understand network performance representative of a specific population,
2. monitoring (sampling) intra- and inter-domain operation, and
3. trouble shooting (on-demand) at the scale of individual users or ISPs.

To support these classes of operation, we define a number of logical components: measurement agents (as the active entities carrying out measurements), measurement servers (as the entities that act as peer points for the measurement agents to perform measurements), one or more controllers (as the instance(s) directing the operations of the measurement agents), and a collector (as a data sink to which the measurement agents upload their results). The operation requires several protocols: between the measurement agents and the controllers to retrieve instructions (schedules, tests to be carried out, servers to be contacted, etc.); the measurement protocols used to execute the tests between the measurement agents and servers; and the upload/collection protocols to store the measurement results. These protocols could be complemented by data formats (for measurement data and metadata) and possibly query formats to access the results database. Finally, mechanisms for software upgrades may be provided to update the measurement agents.

When measurements are carried out not just against measurement servers, but by contacting hosts of service providers (to get a more accurate reading of application performance), we also foresee the necessity of a “do not probe” mechanism by means of which sites can indicate that they do not wish to be measured (conceptually similar to robots.txt for web servers). Other mechanisms may be defined to indicate the willingness of sites to participate in measurements (e.g., using DNS SRV records) as well as to limit the volume of measurement traffic incurred to a given site.

4 Using a Global Measurement Framework

Al Morton (AT&T), Henning Schulzrinne (Columbia University, FCC), Andrea Soppera (BT), Fabian Bustamante (Northwestern University) introduced the topic of use cases for a global measurement framework. We originally considered three use cases defining who the measurement results are targeted at:

1. The operator use case, in which operators use measurements for monitoring and optimizing their networks;
2. the regulator use cases, in which a government entity wants to oversee that the operators fulfill their obligations and do not overclaim the services they are offering;
3. the end user use case, in which measurements assist the end users, e.g., in validating the services they are obtaining and, (in conjunction with operator support) in resolving access or performance problems.

One special case related to end users are application designers, whose applications could learn from measurements about the expected performance (or changes therein) and react accordingly at runtime.

Across the use cases, the “target” for measurement results may differ. On the one hand, there is a technical audience (engineers, researchers, etc.) interested in improving the (cost

effectiveness of) network services and performing trouble shooting when needed. On the other hand, we have regulators and (company) lawyers and further less technical people who also need or want to understand and work with results from network measurements, e.g., to ensure compliance with government regulations, compare networks, etc. While probably all metrics can be gamed in one way or another, the risk of being caught (since users, peers, competitors, and third parties are monitoring as well) is substantial, so that there is little incentive for cheating – metrics won't need to be protected from this perspective. Nevertheless, it would be nice if we could define metrics in a way that if an operator attempts gaming them, this would result in performance (or other) benefits for the user.

Two working groups were formed: 1) One use case covering trouble isolation for operators (which also covers elements of the end user use case) and 2) one addressing data analysis in general.

4.1 Use Case: Service Provider Trouble Isolation

End users carry out measurements because they do care about their network performance – this is reflected in speedtest.net having seen more than 5bn measurements. There are many reasons for this, including: a user's experience may be unsatisfactory; a user may have a new service subscription (ISP or content) and wants to see if it lives up to expectations; a user may have bought new equipment and wants to see its (improved) performance; or a user may carry out tests, possibly as a byproduct of another activity.

However, carrying out such user-invoked measurements using some of the most prominent test platforms may actually not help very much: the user only makes a single measurement point at a time, without calibration to a baseline as discussed above. Performing a ping-based latency (RTT) measurement and performing then an end-to-end file transfer to a point of the measurement system fails to localize the issues (they may not even show any issues if the problem is in a network segment not traversed by the test) and are of unknown accuracy. While ISPs have carefully managed networks, some segments of the path are not managed at all: this includes especially the user's home network. This would require separate measurements, especially when WLANs are involved, given that the WLAN channels used overlap in many buildings with unpredictable performance impact.

If we are able to deploy measurement points at the edge of the network and coordinate measurements from endpoints or home network devices with such embedded measurement point deployments, we can help customers isolate whether the problems is in their home network or the access network. Carrying out measurements may influence future quality of service and quality of experience for users and can yield a positive experience for the users and improve satisfaction with the subscribed service.

There is a tension concerning privacy: the more data is available (instantaneous and historic data) about a user, the more effective trouble shooting can be carried out; yet, at the same time, there is a legitimate desire to maximize user privacy, e.g., by minimizing the amount of data collected and stored. In some scenarios there may also be a tension with business sensitivity.

4.2 Data Analysis

This group addressed the mechanics of the data processing required for data analysis. First of all, generic cloud computing services (by third parties) should not be used because moving around all the large (and constantly growing) volumes of data may be hard because there are issues of trust with the cloud service provider (and well as the network), among other reasons. The consequence is that entities running measurement platforms build their own (post)processing cloud. We look at three case studies.

Akamai collects data for billing purposes as well as for optimizations. For billing, they collect data in quasi real-time (1 min delay), moving the data from the caches to their data center, perform aggregation using Hadoop (HDFS, hbase), and keep the data for diagnosis for two weeks and those data needed for legal purposes for two months. For optimization, DNS to cache allocations are recomputed once per minute based upon the observed performance. RIPE collects data for statistics purposes and long-term observations. They also use Hadoop (HDFS, hbase). Data collected is aggregated, the volume is reduced, and the then preprocessed to make the data sets accessible to tools such as R.

Ftw and Polito collect data from passive measurements, so the resulting data volume gets really large. They store data in SQL with a custom data warehouse solution or Hadoop, respectively, with customized post-processing. In all cases, the collected data is used for reporting purposes (structured repetitive tasks), data mining (more ad-hoc and relying on individual ingenuity for analysis). All have in common that they (have to) use custom-developed processing and evaluation solutions. What is missing is a common toolset / platform that offers a basic set of functionality applicable for the needs across the different platform described above. This also extends towards visualization and to a framework (and formats) for sharing data.

5 Impressions and Next Steps

This Dagstuhl seminar saw 2.5 days (and evenings) of lively and extensive discussions among the participants. The different stakeholders were well represented and also the mix of academia and industry was just right. Sharing perspectives and experience from their respective viewpoints was extremely valuable. We clearly made progress in understanding the issues at hand and important steps to be taken, which we will also feed into the discussion of the different working groups at the IETF. We also foresee work on a joint scientific publication documenting the insights gained in this seminar. Finally, the participants expressed strong interest in continuing our discussions as a follow-up seminar in the future.

Acknowledgements. Two EC FP7 research projects, Leone and mPlane, kindly supported the social event of the seminar.

Participants

- Saba Ahsan
Aalto University, FI
- Vaibhav Bajpai
Jacobs Universität – Bremen, DE
- Arthur W. Berger
Akamai Technologies –
Cambridge, US
- Ernst Biersack
EURECOM – Biot, FR
- Trevor Burbridge
British Telecom R&D –
Ipswich, GB
- Fabian E. Bustamante
Northwestern University –
Evanston, US
- Pedro Casas
FZ Telekommunikation Wien, AT
- Benoit Claise
CISCO Systems Belgium, BE
- Sam Crawford
SamKnows Ltd. – London, GB
- Philip Eardley
British Telecom R&D –
Ipswich, GB
- Daniel Karrenberg
RIPE NCC – Amsterdam, NL
- Mirja Kühlewind
Universität Stuttgart, DE
- Abdelkader Lahmadi
INRIA – Nancy – Grand Est, FR
- Jukka Manner
Aalto University, FI
- Marco Mellia
Polytechnic Univ. of Torino, IT
- Al Morton
AT&T – Middletown, US
- Jörg Ott
Aalto University, FI
- Fabio Ricciato
AIT – Wien, AT
- Dario Rossi
Télécom Paris Tech, FR
- Ramin Sadre
Aalborg University, DK
- Jürgen Schönwälder
Jacobs Universität – Bremen, DE
- Henning Schulzrinne
Columbia Univ. – New York, US
- Andrea Soppera
British Telecom R&D –
Ipswich, GB
- Anna Sperotto
University of Twente, NL
- Burkhard Stiller
Universität Zürich, CH
- Tivadar Szemethy
Netvisor – Budapest, HU
- Brian Trammell
ETH Zürich, CH)



Unleashing Operational Process Mining

Edited by

Rafael Accorsi¹, Ernesto Damiani², and Wil van der Aalst³

1 Universität Freiburg, DE, rafael.accorsi@iig.uni-freiburg.de

2 Università degli Studi di Milano – Crema, IT, ernesto.damiani@unimi.it

3 Eindhoven University of Technology, NL, w.m.p.v.d.aalst@tue.nl

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 13481 “Unleashing Operational Process Mining”. Process mining is a young research discipline connecting computational intelligence and data mining on the one hand and process modeling and analysis on the other hand. The goal of process mining is to discover, monitor, diagnose and improve real processes by extracting knowledge from event logs readily available in today’s information systems. Process mining bridges the gap between data mining and business process modeling and analysis. The seminar that took place November 2013 was the first in its kind. About 50 process mining experts joined forces to discuss the main process mining challenges and present cutting edge results. This report aims to describe the presentations, discussions, and findings.

Seminar 25.–29. November, 2013 – www.dagstuhl.de/13481

1998 ACM Subject Classification H.2.8 Database Applications (Data mining), H.2.1 Logical Design (Data models), K.6 Management of Computing and Information Systems

Keywords and phrases Process mining, Big data, Conformance checking

Digital Object Identifier 10.4230/DagRep.3.11.154

1 Executive Summary

Rafael Accorsi

Ernesto Damiani

Wil van der Aalst

License © Creative Commons BY 3.0 Unported license
© Rafael Accorsi, Ernesto Damiani, and Wil van der Aalst

Society shifted from being predominantly “analog” to “digital” in just a few years. This has had an incredible impact on the way we do business and communicate. Gartner uses the phrase “The Nexus of Forces” to refer to the convergence and mutual reinforcement of four interdependent trends: social, mobile, cloud, and information. The term “Big Data” is often used to refer to the incredible growth of data in recent years. However, the ultimate goal is not to collect more data, but to turn data into real value. This means that data should be used to improve existing products, processes and services, or enable new ones.

Event data are the most important source of information. Events may take place inside a machine (e.g., an X-ray machine or baggage handling system), inside an enterprise information system (e.g., an order placed by a customer), inside a hospital (e.g., the analysis of a blood sample), inside a social network (e.g., exchanging e-mails or twitter messages), inside a transportation system (e.g., checking in, buying a ticket, or passing through a toll booth), etc.



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

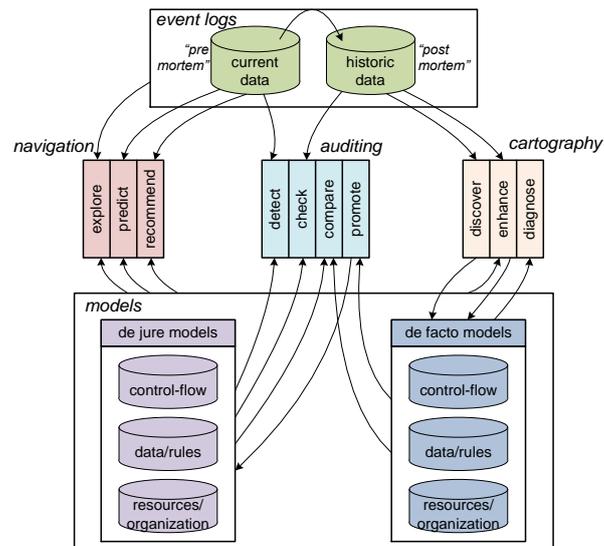
Unleashing Operational Process Mining, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 154–192

Editors: Rafael Accorsi, Ernesto Damiani, and Wil van der Aalst



DAGSTUHL
REPORTS

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** Overview of the different process mining tasks (taken from “Process Mining: Discovery, Conformance and Enhancement of Business Processes”).

Process mining aims to *discover, monitor and improve real processes by extracting knowledge from event logs* readily available in today’s information systems¹. The starting point for process mining is an *event log*. Each event in such a log refers to an *activity* (i.e., a well-defined step in some process) and is related to a particular *case* (i.e., a *process instance*). The events belonging to a case are *ordered* and can be seen as one “run” of the process. Event logs may store additional information about events. In fact, whenever possible, process mining techniques use extra information such as the *resource* (i.e., person or device) executing or initiating the activity, the *timestamp* of the event, or *data elements* recorded with the event (e.g., the size of an order).

Event logs can be used to conduct three types of process mining. The first type of process mining is *discovery*. A discovery technique takes an event log and produces a model without using any a-priori information. Process discovery is the most prominent process mining technique. For many organizations it is surprising to see that existing techniques are indeed able to discover real processes merely based on example behaviors stored in event logs. The second type of process mining is *conformance*. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The third type of process mining is *enhancement*. Here, the idea is to extend or improve an existing process model thereby using information about the actual process recorded in some event log. Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-priori model. For instance, by using timestamps in the event log one can extend the model to show bottlenecks, service levels, and throughput times.

Process mining algorithms have been implemented in various academic and commercial

¹ Process Mining: Discovery, Conformance and Enhancement of Business Processes by W. M. P. van der Aalst, Springer Verlag, 2011 (ISBN 978-3-642-19344-6).

systems. The corresponding tools are being increasingly relevant in industry and have proven to be essential means to meet business goals. ProM is the de facto standard platform for process mining in the academic world. Examples of commercial tools are Disco (Fluxicon), Perceptive Process Mining (before Futura Reflect and BPM|one), QPR ProcessAnalyzer, ARIS Process Performance Manager, Celonis Discovery, Interstage Process Discovery (Fujitsu), Discovery Analyst (StereoLOGIC), and XMAalyzer (XMPro). Representatives of ProM community and the first three commercial vendors participated in Dagstuhl Seminar 13481 “Unleashing Operational Process Mining”.

The Dagstuhl Seminar was co-organized with the *IEEE Task Force on Process Mining* (see <http://www.win.tue.nl/ieeetfpm/>). The goal of this Task Force is to promote the research, development, education and understanding of process mining. Sixty organizations and over one hundred experts have joined forces in the IEEE Task Force on Process Mining.

Next to some introductory talks (e.g., an overview of the process mining field by Wil van der Aalst), 31 talks were given by the participants. The talks covered the entire process mining spectrum, including:

- from theory to applications,
- from methodological to tool-oriented,
- from data quality to new analysis techniques,
- from big data to semi-structured data,
- from discovery to conformance,
- from health-care to security, and
- from off-line to online.

The abstracts of all talks are included in this report.

It was remarkable to see that all participants (including the academics) were very motivated to solve real-life problems and considered increasing the adoption of process mining as one of the key priorities, thereby justifying the title and spirit of the seminar, namely “Unleashing the Power of Process Mining”. This does not imply that there are not many foundational research challenges. For example, the increasing amounts of event data are creating many new challenges and new questions have emerged. Such issues were discussed both during the sessions and on informal meetings during the breaks and at the evening.

Half of the program was devoted to discussions on a set of predefined themes. These topics were extracted based on a questionnaire filled out by all participants before the seminar.

1. Process mining of multi-perspective models (Chair: Akhil Kumar)
2. Data quality and data preparation (Chair: Frank van Geffen)
3. Process discovery: Playing with the representational bias (Josep Carmona)
4. Evaluation of process mining algorithms: benchmark data sets and conformance metrics (Chair: Boudewijn van Dongen)
5. Advanced topics in process discovery: on-the-fly and distributed process discovery (Chair: Alessandro Sperduti)
6. Process mining and Big Data (Chair: Marcello Leida)
7. Process mining in Healthcare (Chair: Pnina Soffer)
8. Security and privacy issues in large process data sets (Chair: Simon Foley, replacing Günter Müller)
9. Conformance checking for security, compliance and auditing (Chair: Massimiliano De Leoni, replacing Marco Montali)
10. How to sell process mining? (Chair: Anne Rozinat)

11. What is the ideal tool for an expert user? (Chair: Benoit Depaire)
12. What is the ideal tool for a casual business user? (Chair: Teemu Lehto)

Summaries of all discussions are included in this report. The chairs did an excellent job in guiding the discussions. After the each discussion participants had a better understanding of the challenges that process mining is facing. This definitely include many research challenges, but also challenges related to boosting the adoption of process mining in industry.

The social program was rich and vivid, including an excursion to Trier's Christmas market, a night walk to ruins, table football, table tennis, and late night discussions.

Next to this report, a tangible output of the seminar is a special issue of IEEE Transactions on Services Computing based on the seminar. This special issue has the title "Processes Meet Big Data" and will be based on contributions from participants of this seminar (also open to others). This special issue of IEEE Transaction on Service-Oriented Computing is intended to create an international forum for presenting innovative developments of process monitoring, analysis and mining over service-oriented architectures, aimed at handling "big logs" and use them effectively for discovery, dash-boarding and mining. The ultimate objective is to identify the promising research avenues, report the main results and promote the visibility and relevance of this area.

Overall, the seminar was very successful. Most participants encouraged the organizers to organize another Dagstuhl Seminar on process mining. Several suggestions were given for such a future seminar, e.g., providing event logs for competitions and complementary types of analysis before or during the seminar. These recommendations were subject of the discussion sessions, whose summaries can be found below.

2 Table of Contents

Executive Summary

<i>Rafael Accorsi, Ernesto Damiani, and Wil van der Aalst</i>	154
-------------------------------------------------------------------------	-----

Overview of Talks

Knowledge and Business Intelligence Technologies in Cross-Enterprise Environments (KITE.it) <i>Antonio Caforio</i>	161
To Unleash, or not to Unleash, that is the Question! <i>Josep Carmona</i>	161
Mining Collaboration in Business Process <i>Paolo Ceravolo</i>	161
Agile Process Mining? <i>Jonathan Cook</i>	162
Sesar Lab Activities <i>Ernesto Damiani</i>	163
Building Better Simulation Models with Process Mining <i>Benoit Depaire</i>	163
Mining the Unknown Security Frontier from System Logs <i>Simon N. Foley</i>	163
Beyond Tasks and Gateways: Towards Rich BPMN Process Mining <i>Luciano Garcia-Banuelos</i>	164
Working with BPMN in ProM <i>Anna A. Kalenkova</i>	164
Benchmarking Process Mining Algorithms on Noisy Data: Does Log Sanitization Help? <i>Akhil Kumar</i>	165
Analytics for Case Management and other Semi-Structured Environments <i>Geetika T. Lakshmanan</i>	165
Continuous Data-driven Business Process Improvement for SAP Order To Cash process <i>Teemu Lehto</i>	165
QPR ProcessAnalyzer (Tool Demo) <i>Teemu Lehto</i>	166
Big Data Techniques for Process Monitoring <i>Marcello Leida</i>	166
Process Mining and BigData <i>Marcello Leida</i>	166
On the Suitability of Process Mining to Produce Current-State Role-based Access Control Models <i>Maria Leitner</i>	167

Declarative Process Mining with ProM <i>Fabrizio Maria Maggi</i>	167
Scientific Workflows within the Process Mining Domain <i>Ronny S. Mans</i>	168
Conformance Analysis of Inventory Processes using Process Mining <i>Zbigniew Paszkiewicz</i>	168
Predictive Security Analysis@Runtime – Lessons Learnt from Adaptation to Industrial Scenarios <i>Roland Rieke</i>	169
Data Collection, Integration, and Cleaning for Process Mining: Reflections on Some Projects <i>Stefanie Rinderle-Ma</i>	169
Disco (Tool Demo) <i>Anne Rozinat</i>	170
DPMine/P Complex Experiment Model Markup Language as Applied to ProM <i>Sergey A. Shershakov</i>	170
A Process Mining-based Analysis of Intentional Noncompliance <i>Pnina Soffer</i>	171
Outpatient Process Analysis with Process Mining: A Case Study <i>Minseok Song</i>	171
PROMPT: Process Mining for Business Process Improvement <i>Alessandro Sperduti</i>	171
SecSy: Security-aware Synthesis of Process Event Logs <i>Thomas Stocker</i>	172
ProM 6.3 (Tool Demo) <i>Eric Verbeek</i>	172
Process Mining in China - Recent Work on Event Quality <i>Jianmin Wang</i>	172
Turning Event Logs into Process Movies: Animating What Has Really Happened <i>Massimiliano de Leoni</i>	173
Introducing Process Mining at Rabobank <i>Frank van Geffen</i>	173
Discussion Sessions	
Discussion Session 1: Process Mining of Multi-Perspective Models	174
Discussion Session 2: Data Quality and Data Preparation	175
Discussion Session 3: Process Discovery: Playing with the Representational Bias	177
Discussion Session 4: Evaluation of Process Mining Algorithms: Benchmark Data Sets and Conformance Metrics	178
Discussion Session 5: Advanced Topics in Process Discovery: On-the-fly and Distributed Process Discovery	179
Discussion Session 6: Process Mining and Big Data	181

160 13481 – Unleashing Operational Process Mining

Discussion Session 7: Process Mining in Health-care 183

Discussion Session 8: Security and Privacy Issues in Large Process Data Sets . . . 184

Discussion Session 9: Conformance Checking for Security, Compliance and Auditing 185

Discussion Session 10: How to Sell Process Mining? 187

Discussion Session 11: What is the Ideal Tool for an Expert User? 188

Discussion Session 12: What is the Ideal Tool for a Casual Business User? 189

Participants 192

3 Overview of Talks

3.1 Knowledge and Business Intelligence Technologies in Cross-Enterprise Environments (KITE.it)

Antonio Caforio (University of Salento, IT)

License © Creative Commons BY 3.0 Unported license
© Antonio Caforio

Main reference L. Fischer (Ed.), “Delivering Competitive Advantage through BPM – Real-World Business Process Management,” Excellence in Practice Series, ISBN 978-0-9849764-5-4, Future Strategies Inc., 2012.

URL <http://futstrat.com/books/CompetitiveAdvantage.php>

In this talk Antonio Caforio described the work ongoing in an Italian industrial research project, Knowledge and business Intelligence Technologies in cross-enterprise Environments (KITE.it), that aims to support the creation and management of processes in the Value Networks (VN). The main project outcomes are methodologies and platforms to enable the alignment of processes with the organizations’ goals in the VN and the measurement of the VN effectiveness. A focus will be made on the Aeroengine Maintenance, Repair and Overhaul (MRO) and its main Overhaul process to understand how process mining can help improve the management of this process.

3.2 To Unleash, or not to Unleash, that is the Question!

Josep Carmona (UPC – Barcelona, ES)

License © Creative Commons BY 3.0 Unported license
© Josep Carmona

In this talk Josep Carmona introduced two very different approaches for unleashing (or not) process mining, that are being developed in my group. The first one, based on the use of portfolio-based algorithm selection techniques, is devoted to guide the application of process mining algorithms by using a recommender system. The second one, totally opposed to the first, aims at providing a process-oriented computing environment for the exploration and creation of process mining algorithms. These two approaches are meant to cover a wide variety of process mining practices and, together with existing frameworks, offer a new perspective to the field.

3.3 Mining Collaboration in Business Process

Paolo Ceravolo (University of Milan, IT)

License © Creative Commons BY 3.0 Unported license
© Paolo Ceravolo

Main reference F. Frati, I. Seeber, “CoPrA: A Tool For Coding and Measuring Communication In Teams,” in Proc. of the 7th IEEE Int’l Conf. on Digital Ecosystems and Technologies (DEST’13), pp. 43–48, IEEE, 2013.

URL <http://dx.doi.org/10.1109/DEST.2013.6611327>

Observing the evolution of several research programs focusing on collaborating communities, we encounter a call for diachronic analysis. It follows that Process Mining can contribute in refining and enriching the next generation of these studies. In whatever way, this implies to understand the research questions that are driving the analysis on collaborative process

to then identify the challenges for evolving Process Mining techniques. In this talk, Paolo Ceravolo looked back on the evolution of some area related to collaborative process and pointing out open issue and interesting research directions.

3.4 Agile Process Mining?

Jonathan Cook (New Mexico State University, US)

License  Creative Commons BY 3.0 Unported license
© Jonathan Cook

Joint work of Cook, Jonathan; Bani-Hani, Imad

Agile software development methodologies are in some sense a reaction to overly prescriptive development processes, and to a large extent desired to throw out a strict process model and move the team collaboration and project management off the computer and make it human centric. Thus we see management tools such as white boards with sticky notes being used to manage the project. Can such a process be mined effectively? We argue yes, because agile methodologies still embody practices that should result in some regular, observable patterns of behavior or at least constraints over what activities take place when. For some examples, test driven design should show itself in the activity of creating a test case (or more) before a feature is implemented; time-boxed iteration should show itself in very regular release tagging; continuous integration should show itself in regular feature merges into the main build; and refactoring should show itself in particular code edit patterns or commit messages. Knowing the process (or practices) that a project team is performing will help them in assessing their own agility and potentially show them areas where they could improve. Beyond the set of recommended practices, agile processes should be flexible; thus very closed-form control-flow process mining algorithms should probably not work well on an agile process (and if they do, it may not be very agile). Open model mining algorithms such as DeclareMiner, which infers particular LTL rule patterns, should be much more suitable for agile process mining. Other aspects of process mining, such as role mining, organizational structure mining, and social network mining may help in agile process mining; for example, an agile team may have a goal to be as interactive and collaborative as possible, sharing duties equally, but in practice they may slip into very specific roles without realizing it. Finally, before pursuing agile process mining, we are creating qualitative methods (e.g., a questionnaire) for measuring the agility of a process, and have defined five dimensions over which to measure agility: team (level of interaction and collaboration practices); customer (level of customer involvement); iteration (level of true iterative practices); testing (level of test-centric practices); and design (level of ongoing refactoring effort). In this talk, Jon Cook asked 3–6 questions in each area and then create a radial chart showing the level of agility for each dimension and giving a visualization of overall agility.

3.5 Sesar Lab Activities

Ernesto Damiani (Università degli Studi di Milano – Crema, IT)

License © Creative Commons BY 3.0 Unported license
© Ernesto Damiani

Secure Service-oriented Architectures Research (SESAR) Lab within the Computer Science Department of the Università degli Studi di Milano. The research activities are mainly focused on the following subjects ranging from Service-oriented Architectures to Knowledge Management over Open Source Development Paradigms and Security. The staff is composed by full time Professors, Researchers, Post-Docs, PhDs and Research Collaborators. The research activities are carried out in collaboration with Italian and European partners, within national and international research projects and agreements with enterprises. The Lab offers to University students the chance to carry out degree theses, stages for the acknowledgment of university credits, and the opportunity to participate to the activities carried out in each research project, achieving experiences for the future work activities.

3.6 Building Better Simulation Models with Process Mining

Benoit Depaire (Hasselt University – Diepenbeek, BE)

License © Creative Commons BY 3.0 Unported license
© Benoit Depaire

Joint work of Depaire, Benoit; Martin, Niels; Caris, An

Simulation models are a useful tool to run complex what-if scenarios and to make informed decisions. However, these simulation models are often constructed in a highly subjective way (through interviews, documents how processes should be executed and guesstimates on simulation parameters). In this talk Benoit Depaire argued that process mining holds the tools and potential to construct more reliable simulation models. We presented a SWOT analysis of business process simulation based on the current state of the art in literature, presented a framework how simulation and process mining could be linked together and identified different challenges where the process mining community should focus on.

3.7 Mining the Unknown Security Frontier from System Logs

Simon N. Foley (University College Cork, IE)

License © Creative Commons BY 3.0 Unported license
© Simon N. Foley

Joint work of Foley, Simon N.; Pieczul, Olgierd

The scale and complexity of modern computer systems has meant that it is becoming increasingly difficult and expensive to formulate effective security policies and to deploy efficacious security controls. As a consequence, security compliance tends to focus on those activities perceived to be critical, with an assumption that the other activities, known or unknown, are not significant. However, often it is these side-activities that can lead to a security compromise of the system. While security controls provide monitoring and enforcement of the critical activities related to the security policy, effectively, little is known about the nature of the other activities. Our preliminary results show that such activities can

be modeled and do exist in real-world systems. In this talk, Simon Foley demonstrated how process mining techniques can be explored in practice to discover and check for perturbations to these activities in system logs.

3.8 Beyond Tasks and Gateways: Towards Rich BPMN Process Mining

Luciano Garcia-Banuelos (University of Tartu, EE)

License  Creative Commons BY 3.0 Unported license
© Luciano Garcia-Banuelos

During the last decade, process mining techniques have reached a certain level of sophistication and maturity, evidenced by the availability of a range of functional academic prototypes and commercial tools in the field. In parallel to these developments, BPMN has emerged as a widely adopted standard for modeling and analyzing business processes. BPMN offers a rich set of constructs for modeling business processes in a structured way, including sub-processes with interrupting and non-interrupting boundary events and multi-instance activities as well as a comprehensive set of event types. Surprisingly though, the bulk of research in process mining in general, and automated process model discovery in particular, has focused on the problem of discovering process models consisting purely of tasks and control-flow dependencies (in essence: tasks and some types of gateways). In this talk, Luciano Garcia-Banuelos presented his initial work on automated discovery of rich BPMN process models, meaning process models that make use of the BPMN notation beyond its “task and gateways” subset. He discussed initial achievements and key challenges he had identified so far.

3.9 Working with BPMN in ProM

Anna A. Kalenkova (NRU Higher School of Economics – Moscow, RU)

License  Creative Commons BY 3.0 Unported license
© Anna A. Kalenkova

ProM is a tool for implementing and integrating process mining algorithms within a standard environment. ProM plugins support plenty of different process model formats, among them are Petri nets, transition systems, casual nets, fuzzy models and others which are widely used by researchers. But at the same time it might be rather difficult for an inexperienced user (or for an external customer) to estimate the result of applying process mining techniques and understand the semantics of process models. This indicates that there is a need for ProM to support commonly known process modeling standards also. BPMN (Business Process Modeling Notation) is a process modeling and executing notation understandable by a wide audience of analytics and developers. Representing process models in this standard way will give an ability to bridge the gap between ProM and variety of process modeling tools. Also BPMN gives a holistic view on the process model: BPMN diagrams could be enhanced with roles, interactions, timers, conformance/performance info, etc. In her talk, Anna Kalenkova gave an overview of ProM functionality related to BPMN. Import/export capabilities and internal BPMN meta-model were discussed. Also the plugins which implement conversions from different formalisms to BPMN and vice-versa were considered.

3.10 Benchmarking Process Mining Algorithms on Noisy Data: Does Log Sanitization Help?

Akhil Kumar (Pennsylvania State University, US)

License  Creative Commons BY 3.0 Unported license
© Akhil Kumar

Akhil Kumar proposed a technique to sanitize noisy logs by first building a classifier on a subset of the log and applying the classifier rules to remove noisy traces from the log. The technique is evaluated on synthetic logs from six benchmark models of increasing complexity on both behavioral and structural recall and precision metrics. The results show that mined models thus produced from sanitized log are superior on the evaluation metrics. They show better fidelity to the reference models and are more compact. The rule based approach generalizes to any noise pattern. The rules can be explained and modified.

3.11 Analytics for Case Management and other Semi-Structured Environments

Geetika T. Lakshmanan (IBM TJ Watson Research Center – Cambridge, US)

License  Creative Commons BY 3.0 Unported license
© Geetika T. Lakshmanan

There is considerable scope for both predictive and descriptive analytics in case management and other semi-structured environments. Predictive analytics could provide guidance to a case worker handling a current case instance on the likelihood of a future task occurrence or attribute value. By training a classifier such as a decision tree on a set of completed case execution traces, the classifier can be used to make predictions about the likelihood of occurrence of a task execution or predict the value of a continuous variable in the case such as time for a currently running case instance. Descriptive analytics could be applied to provide insight about correlations and patterns derived from historically completed instances of a case. In order to be applied in a real world setting, these analytics require solving an array of challenges. In addition to providing easily consumable results, these analytics have to be highly confident of the predictions and correlations they compute. In her talk, Geetika Lakshmanan provided an overview of the challenges of applying predictive and descriptive analytics to case management and other real world settings.

3.12 Continuous Data-driven Business Process Improvement for SAP Order To Cash process

Teemu Lehto (QPR Software – Helsinki, FI)

License  Creative Commons BY 3.0 Unported license
© Teemu Lehto

This talk was a business-driven case study for sharing experiences (1) SAP is the leading ERP system globally measured by revenue. (2) SAP creates great quality records for process mining purposes. (3) Order to cash is a critical importance business process for organizations. (4) Order to cash is not as systematic or optimized as one could think of.

3.13 QPR ProcessAnalyzer (Tool Demo)

Teemu Lehto (QPR Software – Helsinki, FI)

License  Creative Commons BY 3.0 Unported license
© Teemu Lehto

This talk is a demo of a commercial Process Mining tool QPR ProcessAnalyzer.

3.14 Big Data Techniques for Process Monitoring

Marcello Leida (Khalifa University – Abu Dhabi, AE)

License  Creative Commons BY 3.0 Unported license
© Marcello Leida

Joint work of Leida, Marcello, Andrej Chu, Basim Majeed

Main reference M. Leida, A. Chu, “Distributed SPARQL Query Answering over RDF Data Streams,” in Proc. of 2013 IEEE Int’l Congress on Big Data (BigData’13), pp. 369–378, IEEE, 2013.

URL <http://dx.doi.org/10.1109/BigData.Congress.2013.56>

Modern Business process analysis requires an extremely flexible data model and a platform able to minimize response times as much as possible. In order to efficiently analyze a large amount of data, this talk illustrated novel technologies that rely on an improved data model supported by a grid infrastructure, allowing storing the data in-memory across many grid nodes and distributing the workload, avoiding the bottleneck represented by constantly querying a traditional database. Both process data and domain knowledge are represented using standard metadata formats: process logs are stored as RDF triples referring to company specific activities. The data collected by the process log monitor is translated to a continuous flow of triples that capture the status of the processes. This continuous flow of information can be accessed through the SPARQL query language used to extract and analyze process execution data. Although the query engine has been developed as part of a Business Process Monitoring platform, it is a general purpose engine that can be used in any system that requires scalable analysis of semantic data. The system presented has some unique features such as grid-based infrastructure, extreme scalability, efficient real-time query answering and an on the fly access control layer that were presented in detail during the talk.

3.15 Process Mining and BigData

Marcello Leida (Khalifa University – Abu Dhabi, AE)

License  Creative Commons BY 3.0 Unported license
© Marcello Leida

This discussion session focused on the various big data technologies and how can they be applied to process mining area.

3.16 On the Suitability of Process Mining to Produce Current-State Role-based Access Control Models

Maria Leitner (Universität Wien, AT)

License © Creative Commons BY 3.0 Unported license
© Maria Leitner

Joint work of Leitner, Maria; Rinderle-Ma, Stefanie

Main reference M. Leitner, "Delta analysis of role-based access control models," in Proc. of the 14th Int'l Conf. on Computer Aided Systems Theory (EUROCAST'13), Part I, LNCS, Vol. 8111, pp. 507–514, Springer, 2013.

URL http://dx.doi.org/10.1007/978-3-642-53856-8_64

Role-based access control (RBAC) is the de facto standard for access control in process-aware information systems. With existing techniques in organizational mining, we can adapt these to derive not only organizational models but also RBAC models. In a case study, we evaluated role derivation, role hierarchy mining, organizational mining, and staff assignment mining on the suitability to derive RBAC models. We compared the derived models to the original and evaluated the results with quantitative measures. Furthermore, we adapted delta analysis to the RBAC domain to investigate the similarity of RBAC models and to analyze differences between the models. As an example, we analyzed the structural similarity using error correcting graph matching. With this approach, we can not only identify RBAC misconfiguration but also detect violations of the original RBAC policy.

References

- 1 Maria Leitner, Anne Baumgrass, Sigrid Schefer-Wenzl, Stefanie Rinderle-Ma, and Mark Strembeck: A Case Study on the Suitability of Process Mining to Produce Current-State RBAC Models. Business Process Management Workshops. LNBIP. Springer, pp. 719–724 (2013)
- 2 Maria Leitner: Delta Analysis of Role-based Access Control Models: Proceedings of the 14th International Conference on Computer Aided Systems Theory (EUROCAST 2013). LNCS. Springer, pp. 507–514 (2013) (in press)

3.17 Declarative Process Mining with ProM

Fabrizio Maria Maggi (University of Tartu, EE)

License © Creative Commons BY 3.0 Unported license
© Fabrizio Maria Maggi

The increasing availability of event data recorded by contemporary information systems makes process mining a valuable instrument to improve and support business processes. Starting point for process mining is an event log. Typically, three types of process mining can be distinguished: (a) process discovery (learning a model from example traces in an event log), (b) conformance checking (comparing the observed behavior in the event log with the modeled behavior), and (c) model enhancement (extending models based on additional information in the event logs, e.g., to highlight bottlenecks). Existing process mining techniques mainly use procedural process modeling languages for describing the business processes under examination. However, these languages are suitable to be used in stable environment where process executions are highly predictable. In turbulent environments, where process executions involve multiple alternatives, process models tend to be complex and difficult to understand. In this talk, Fabrizio Maggi introduced a new family of process mining techniques based on declarative languages. These techniques are very suitable to be

used for analyzing less structured business processes working in environments where high flexibility is required. These techniques have been implemented in the process mining tool ProM and range from process discovery to models repair and extension, to offline and online conformance checking.

3.18 Scientific Workflows within the Process Mining Domain

Ronny S. Mans (Eindhoven University of Technology, NL)

License  Creative Commons BY 3.0 Unported license
© Ronny S. Mans

Within the process mining domain there is currently no support for the construction and execution of a workflow which describes all analysis steps and their order, i.e. a scientific workflow. In the tool demo we demonstrated how we have integrated the scientific workflow management system RapidMiner with the process mining framework ProM 6. That is, several interesting workflows, consisting of multiple process mining tasks, will be constructed and executed.

3.19 Conformance Analysis of Inventory Processes using Process Mining

Zbigniew Paszkiewicz (Poznan University of Economics, PL)

License  Creative Commons BY 3.0 Unported license
© Zbigniew Paszkiewicz

Case study: Conformance analysis of inventory processes using process mining Production companies monitor deviations from the assumed procedures to satisfy quality requirements. In his talk, Zbigniew Paszkiewicz showed how process mining contributes to quality management efforts by analysis event logs about inventory operations registered in a warehouse management system. The analyzed company has pointed six aspects to be scrutinized: 1. conforming to model: inventory process instances must follow a pre-defined de jure model; 2. First In First Out policy: products that were produced first must be shipped first; the FIFO rule must be satisfied within particular product families; 3. quality assurance: all the pallets before being shipped to a client must be checked by the quality department; 4. process performance: a particular pallet cannot be stored in the warehouse for more than fourteen days; additional constraints concern the execution time of particular activities related to pallet management; 5. pallet damage handling: a pallet in disrepair must be transported to a special storage area; all the storekeepers are responsible for handling damaged pallets in this way; 6. work distribution: all the shifts should perform an equal amount of work; storekeepers taking pallets from production lines should not be involved in their shipping from the warehouse, and vice-versa. Conformance checking analysis has been performed with both ProM and commercial tools. Unwanted and repeatable parts of inventory processes in their business contexts have been identified with our novel RMV method. Unwanted and repeatable parts are represented as activity patterns which encompass the definition of activities and social relations among process participants. Preliminary results confirm that the RMV method provides useful insights about collaboration among process participants.

3.20 Predictive Security Analysis@Runtime – Lessons Learnt from Adaptation to Industrial Scenarios

Roland Rieke (Fraunhofer SIT – Darmstadt, DE)

License © Creative Commons BY 3.0 Unported license
© Roland Rieke

Joint work of Rieke, Roland; Repp, Jürgen; Zhdanova, Maria; Eichler, Jörn

Main reference R. Rieke, J. Repp, M. Zhdanova, J. Eichler, “Monitoring Security Compliance of Critical Processes,” in Proc. of the 22th Euromicro Int’l Conf. on Parallel, Distributed and Network-Based Processing (PDP’14), IEEE CS, to appear.

The Internet today provides the environment for novel applications and processes which may evolve way beyond pre-planned scope and purpose. Security analysis is growing in complexity with the increase in functionality, connectivity, and dynamics of current electronic business processes. Technical processes within critical infrastructures also have to cope with these developments. To tackle the complexity of the security analysis, the application of models is becoming standard practice. However, model-based support for security analysis is not only needed in pre-operational phases but also during process execution, in order to provide situational security awareness at runtime. This talk given by Roland Rieke presented an approach to support model-based evaluation of the security status of process instances. In particular, challenges with respect to the assessment whether instances of processes violate security policies or might violate them in the near future were addressed. The approach is based on operational formal models derived from process specifications and security compliance models derived from high-level security and safety goals. Events from process instances executed by the observed system are filtered for their relevance to the analysis and then mapped to the model of the originating process instance. The applicability of the approach is exemplified utilizing processes from several industrial scenarios. Lessons learnt from the adaptation of the method to the scenarios are addressed. In particular, event model abstraction, process instance identification, semi-automatic model mining, and cross process instance reasoning is discussed. Furthermore, the need for a method to derive measurement requirements from security and dependability goals is motivated and a meta model aiming at an integrated security strategy management is presented.

3.21 Data Collection, Integration, and Cleaning for Process Mining: Reflections on Some Projects

Stefanie Rinderle-Ma (Universität Wien, AT)

License © Creative Commons BY 3.0 Unported license
© Stefanie Rinderle-Ma

Joint work of Rinderle-Ma, Stefanie; Ly, Linh Thao; Mangler, Jürgen; Indiono, Conrad; Dunkl, Reinhold; Kriglstein, Simone; Wallner, Günter; Binder, Michael; Dorda, Wolfgang; Duftschmid, Georg; Fröschl, Karl Anton; Gall, Walter; Grossmann, Wilfried; Harmankaya, Kaan; Hronsky, Milan; Rinner, Christoph; Weber, Stefanie

Main reference L. Thao Ly, C. Indiono, J. Mangler, S. Rinderle-Ma, “Data Transformation and Semantic Log Purging for Process Mining,” in Proc. of the 24th Int’l Conf. on Advanced Information Systems Engineering (CAiSE’12), LNCS, Vol. 7328, pp. 238–253, Springer, 2012.

URL http://dx.doi.org/10.1007/978-3-642-31095-9_16/

In this talk, Stefanie Rinderle-Ma highlighted some process mining challenges from her own projects. The first project is EBMC2 which is a joint work between University of Vienna and Medical University of Vienna on patient treatment processes in skin cancer. The goal of the project is to discover the actual treatment processes and compare them with skin cancer

guidelines in order to analyze possible deviations. Though several data sources are available several data integration and quality problems occur, e.g., with respect to activity granularity and time. The second project is on higher education processes (HEP) where we tried to mine reference processes based on semantic log purging. Finally, some results on visualizing process difference graph including instance traffic are presented.

References

- 1 Michael Binder, Wolfgang Dorda, Georg Duftschmid, Reinhold Dunkl, Karl Anton Fröschl, Walter Gall, Wilfried Grossmann, Kaan Harmankaya, Milan Hronsky, Stefanie Rinderle-Ma, Christoph Rinner, Stefanie Weber: On Analyzing Process Compliance in Skin Cancer Treatment: An Experience Report from the Evidence-Based Medical Compliance Cluster (EBMC2). Int'l Conf on Advanced Information Systems Engineering (CAiSE'12), pp. 398–413 (2012)
- 2 Linh Thao Ly, Conrad Indiono, Jürgen Mangler, Stefanie Rinderle-Ma: Data Transformation and Semantic Log Purging for Process Mining. Int'l Conf on Advanced Information Systems Engineering (CAiSE'12), pp. 238–253 (2012)
- 3 Simone Kriglstein, Günter Wallner, Stefanie Rinderle-Ma: A Visualization Approach for Difference Analysis of Process Models and Instance Traffic. Int'l Conf on Business Process Management (BPM'13), pp. 219–226 (2013)

3.22 Disco (Tool Demo)

Anne Rozinat (*Fluxicon Process Laboratories, NL*)

License  Creative Commons BY 3.0 Unported license
© Anne Rozinat

This talk presented Disco, a professional tool for process mining practitioners.

3.23 DPMine/P Complex Experiment Model Markup Language as Applied to ProM

Sergey A. Shershakov (*NRU Higher School of Economics – Moscow, RU*)

License  Creative Commons BY 3.0 Unported license
© Sergey A. Shershakov

In his talk, Sergey Shershakov considered DPMine/P, a new language for modeling domain-specific Process Mining experiments, and tool support for this language based on ProM platform. The language under development aims at the unification of the separate phases of an experiment into a single sequence, that is an experiment model, support of looping constructs and other execution threads controls, provision of a clear but flexible (and, what is important, expandable) semantics. DPMine/P language is considered at the level of ProM tool as a set of plug-ins and data objects (which are the input and output data for the plug-ins). A description of some modules and examples of their use is provided.

3.24 A Process Mining-based Analysis of Intentional Noncompliance

Pnina Soffer (University of Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Pnina Soffer

Business process workarounds are specific forms of incompliant behavior, where employees intentionally decide to deviate from the required procedures although they are aware of them. Detecting and understanding the workarounds performed can guide organizations in redesigning and improving their processes and support systems. In this talk, Pnina Soffer presents her work on building specific types of workarounds found in practice, and defining corresponding log patterns for detecting them by process mining. Pnina analyzed logs of 5 real-life processes and find correlations between the frequency of specific workaround types and properties of the processes and of specific activities. The analysis results promote the understanding of workaround situations and sources.

3.25 Outpatient Process Analysis with Process Mining: A Case Study

Minseok Song (UNIST, KR)

License  Creative Commons BY 3.0 Unported license
© Minseok Song

In the talk of Minseok Song, a case study with a real life log from a hospital in Korea is explained. Based on the outpatients, event log in the hospital, he derived the process model and compared it with the standard model in the hospital. In addition, he conducted performance analysis to make a simulation model and analyzed the process patterns according to patient types. According to the result of comparing the event log and their standard process model, the matching rate was as 89.01%. That is, they relatively well understood workflows of outpatients and the process was well-managed by the hospital. Using the performance analysis result, he generated the simulation model. The simulation shows that the 10% increase of patients makes the largest change in consultation waiting time. Thus, he recommended less than 10% of increase. He extracted the process models and analyzed the process patterns according to patient types. The most frequent pattern of each patient type was discovered. The patterns are used to build a smart guidance app in the ubiquitous healthcare system in the hospital. As a future work, he will analyze more processes such as call clinical pathways, payment processes, etc.

3.26 PROMPT: Process Mining for Business Process Improvement

Alessandro Sperduti (University of Padova, IT)

License  Creative Commons BY 3.0 Unported license
© Alessandro Sperduti

This talk presented the PROMPT project and some of the results achieved by the Italian partners. Specifically, I present the basic ideas underpinning: a software for importing data from target information systems; a role mining algorithm; an approach for automatic selection of values for discovery algorithms parameters; a family of algorithms for on-the-fly process discovery. Work in progress is outlined as well.

3.27 SecSy: Security-aware Synthesis of Process Event Logs

Thomas Stocker (Universität Freiburg, DE)

License  Creative Commons BY 3.0 Unported license
© Thomas Stocker

One difficulty at developing mechanisms for business process security monitoring and auditing is the lack of representative, controllably generated test runs to serve as an evaluation basis. SecSy tries to fill this gap by providing tool support for event log synthesis. The novelty is that it considers the activity of an “attacker” able to purposefully infringe security and compliance requirements or simply manipulate the process’ control and data flow, thereby creating deviations of the intended process model. The resulting logs can be readily replayed on a reference monitor, or serve as input for auditing tools based upon, e.g., process mining.

3.28 ProM 6.3 (Tool Demo)

Eric Verbeek (Eindhoven University of Technology, NL)

License  Creative Commons BY 3.0 Unported license
© Eric Verbeek

Process mining has emerged as a way to analyze business processes based on event logs. These events logs need to be extracted from operational systems and can subsequently be used to discover or check the conformance of processes. ProM is a widely used tool for process mining. Earlier versions of ProM were distributed under the CPL license, required a GUI to run, and came with all functionality in a single bundle. As a result, it was not possible to run a mining algorithm from, say, a command line prompt, and we had problems using third-party libraries that came with a conflicting license. ProM 6 overcomes these problems, and ProM 6.3 is the latest version in this line of ProM releases. ProM 6.3 can be downloaded from <http://www.promtools.org/prom6>.

3.29 Process Mining in China - Recent Work on Event Quality

Jianmin Wang (Tsinghua University Beijing, CN)

License  Creative Commons BY 3.0 Unported license
© Jianmin Wang

Business process management has been used in Chinese enterprises widely in last 10 years. Investigating the accumulated event logs will enhance their competition capacities. However, the event quality is often not good enough. In this talk, Jianmin Wang introduced his recent work on event quality. 1) He studied the efficient techniques for recovering missing events. Advanced indexing and pruning techniques based on Petri net unfolding theories are developed to improve the recovery efficiency. 2) A generic pattern based matching framework was proposed, which is compatible with the existing structure based techniques. To improve the matching efficiency, he devised several bounds of matching scores for pruning. 3) An algorithm of mining the non-free choice structure from the dirty log with missing event was also introduced. Finally, the academic research groups in China are emulated and future research directions of our group are presented.

3.30 Turning Event Logs into Process Movies: Animating What Has Really Happened

Massimiliano de Leoni (University of Padova, IT)

License  Creative Commons BY 3.0 Unported license
© Massimiliano de Leoni

Today's information systems log vast amount of data which contains information about the actual execution of business processes. The analysis of this data can provide a solid starting point for business process improvement. This is the realm of process mining, an area which has provided a repertoire of many analysis techniques. Despite the impressive capabilities of existing process mining algorithms, dealing with the abundance of data recorded by contemporary systems and devices remains a challenge. Of particular importance is the capability to guide the meaningful interpretation of this "ocean" of data by process analysts. To this end, insights from the field of visual analytics can be leveraged. The talk discussed an approach where process states are reconstructed from event logs and visualised in succession, leading to an animated history of a process. This approach is customizable in how a process state, partially defined through a collection of activity instances, is visualized: one can select a map and specify a projection of activity instances on this map based on their properties. The approach is implemented as plug-in for process-mining framework ProM. The talk will also show the application to a case study with one of Australia's largest insurance companies: Suncorp.

3.31 Introducing Process Mining at Rabobank

Frank van Geffen (Rabobank – Utrecht, NL)

License  Creative Commons BY 3.0 Unported license
© Frank van Geffen

Our challenge is to match the customer needs of tomorrow. The speed and complexity of today's changes require a different approach to process improvement. Process mining, or automated business process discovery, is a bpm technique that helps in gaining insight in how processes are actually performed, how systems are used and how people work together. Through the explosive growth of data and significant advances in analysis and visualization technology it's possible to unlock valuable process information by analyzing transaction data. The use of automated business process discovery techniques yield new valuable insights. Process analysis done this way becomes fact based, full, for real and fast. Frank van Geffen told about his experience with introducing this new technology at Rabobank. Based on his practical experience, he stated the specific value of this technique for Rabobank. Besides the successes, Frank will also share the pitfalls he encountered and what measures can be taken to circumvent these obstacles.

4 Discussion Sessions

The seminar comprised 12 discussion sessions. Figure 2 depicts their organization and the chairs of each sessions. The following provides a summary of these discussion sessions, as reported by the discussion chairs.

Process mining of multi-perspective models (Chair: Akhil Kumar)	Data quality and data preparation (Chair: Frank van Geffen)
Process discovery: Playing with the representational bias (Josep Carmona)	Evaluation of process mining algorithms: benchmark data sets and conformance metrics (Boudewijn van Dongen)
Advanced topics in process discovery: on-the-fly and distributed process discovery (Alessandro Sperduti)	Process mining and Big Data (Marcello Leida)
Process mining in Healthcare (Pnina Soffer)	Security and privacy issues in large process data sets (Simon Foley)
Conformance checking for security, compliance and auditing (Massimiliano de Leoni)	How to sell process mining? (Anne Rozinat)
What is the ideal tool for an expert user? (Benoit Depaire)	What is the ideal tool for a casual business user? (Teemu Lehto)

■ **Figure 2** Overview of discussions sessions.

4.1 Discussion Session 1: Process Mining of Multi-Perspective Models

The motivation for this group was to discuss whether it would help to explore multiple perspectives in developing process mining algorithms. The group started out by identifying multiple perspectives in addition to the control flow perspective, i.e.:

- Data
- Resource/role/organizational
- Inter-process communication
- Time, costs, risks, energy consumption
- State of a process
- Performance
- Context

The discussion mainly centered on the data and organizational perspectives. We summarize the main issues discussed.

The complexity problem

To deal with the added complexity of multiple perspectives, one could start from a control flow model and enhance it with data related conditions at choice nodes and roles associated with tasks in the model. This may not always work because in some examples of BPMN model discovery it leads to a “spaghetti model.” Yet it was felt that each new perspective potentially adds value. Two approaches discussed are: 1) Treat each perspective as a layer of an onion, where the order of layers would be situation and need dependent; 2) Analyze each perspective separately, and integrate them. However, it was noted that a clean separation may not always be possible.

Representation Problem

While different perspectives can be used for filtering, clustering and alignment, how do we visualize them? An appeal of the control flow model lies in the ease of its visual representation as a graph. Some aspects can be added to this model by means of conditions and rules at gateway nodes, and association of roles with tasks. However, security constraints like binding and separation of duties would be hard to show. There is also need to avoid clutter and give users an ability to select what perspectives they wish to see and to zoom in-out, etc. as is done with maps. Further, multidimensional information could be displayed in 2-D by pairwise cartography.

Log issues

To perform multi-perspective mining the log must include additional data beyond events and timestamps. Thus, the need for additional data in logs was discussed. It was also noted that data can help to discover causality relationships and thus lead to inference of control flow also. The limits of our analysis capability are naturally limited by the information provided by a user in the log. The group felt it would also be interesting to think about extending the XES standard to include event data.

Conclusions

This discussion group sees value in research on multi-perspective process mining. It sees research challenges in the representation and complexity problems. It is felt that multiple perspectives can be analyzed serially or separately based on user questions and data availability. Finally, a need is perceived for user-definable interfaces that allow selection of perspectives, zoom feature, etc. and for extending the XES standard to represent event data in a log.

4.2 Discussion Session 2: Data Quality and Data Preparation

The purpose of the discussion was to gain insights into practical data quality and preparation challenges experienced by the participants. To this end, we first collected typical data challenges from the group and then presented our own challenges, which we had prepared. Frank discussed some challenges based on two real examples from the Rabobank to make it concrete. To summarize the data quality problems/issues regarding the *data quality*, we used an existing framework² to categorize the challenges we had collected in the group and before:

It became apparent during the discussions that missing events and timestamp problems were the most frequently issues mentioned. The discussion session discussed further data quality issues, which for the sake of space will not be described in detail here.

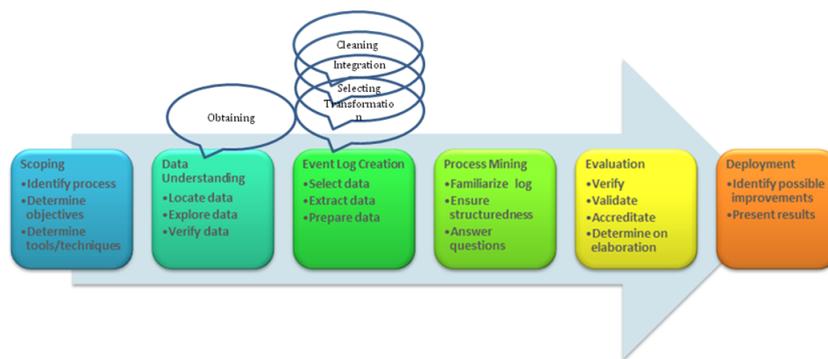
Turning to the *data preparation*, we added the data preparation phase “Obtaining” before “Cleaning”, “Integration”, “Selection”, and “Transformation” as a result of the discussion. To position the discussed activities in the context of a process mining project, we used an existing lifecycle model³ that also illustrates the iterative nature of data preparation, validation, data cleaning, etc. This is depicted in Fig. 4.

² R. P. Jagadeesh Chandra Bose, Ronny S. Mans, Wil M. P. van der Aalst. Wanna Improve Process Mining Results? It's High Time We Consider Data Quality Issues Seriously. BPM Center Report BPM-13-02, BPMcenter.org, 2013

³ T. van der Heijden, Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice. Master Thesis, 2012

	case	event	belongs to	c attribute	position	activity name	timestamp	resource	e attribute
<i>Missing Data</i>	I	V	III			I	II		I
<i>Incorrect Data</i>			I				III		I
<i>Imprecise Data</i>				I			II	I	I
<i>Irrelevant Data</i>		III							
	I		II		III	IV		V	

■ **Figure 3** Data quality categories.



■ **Figure 4** Lifecycle for data preparation.

The complete list of collected data preparation challenges is the following:

- *Obtaining*: Diverse data sources; Very large datasets; Transportation of data; Same tables for different processes; What's the right process / activity scope?; Lack / mismatch between technical (data) documentation and reality; Access to different domain roles / function for understanding.
- *Cleaning*: Validation following individual cases (verifying data quality); Removing duplicates; Correcting date-timestamps; Action codes translation to real names (human-readable, URL to semantic action); What's `case_id` pointing at? (customers / products / documents / complaints / combinations.); Sampling (understanding data); Server times vs. Local times.
- *Integration*: Merging data sources; How to deal with large blobs full of free text?; Large file size, long waiting time; Usage of different separator `<quotes>`; Which merging sequence / order of steps? (when to merge which data / first merge then proces view or visa versa / automation).
- *Selection*: Connecting multiple case IDs to follow end-to-end process; What amount of data is required? (as much as possible, prize, anonymizing, decision criteria); Does the dataset need to be enriched?; Sampling (criteria?).
- *Transformation*: Formatting: activities in columns, you lose loops and assume a pre-specified process is followed; Server times vs. Local times; Which environment is receiving the prepared data? (Disco sets other standards than Rapid miner, Tableau or Click view for example); Working with / without weekends and night-shifts (adjusting timestamps to match the organizations opening and closing times); Multiple `case_id`'s (`customer_id`, `product_id`, `document_id`, `update_id`, `session_id`, `authorization_id`).

The discussion then elicited the following questions and action points:

- *Questions*
 - How to teach checking and communicating data requirements for process mining?
 - Are there tools to help with the understanding of the data properties?
 - What are the consequences of the different data quality levels from the manifesto on certain process analysis questions?
 - When to use sampling and according to which criteria?
 - Could there be a tool to automatically guide through the data preparation process?
 - Organize seminars together with the database community?
- *Actions*
 - Share data sets online.
 - Share best practices.
 - BPI Challenge: compare and summarize the submissions to gain insight in the different approaches

4.3 Discussion Session 3: Process Discovery: Playing with the Representational Bias

The goal of this discussion was to identify the main challenges the representational bias brings into the process discovery discipline. One of the initial discussion run through the proper notion, specially what characterizes a discovery algorithm in terms of the implicit bias it has.

Right after discussing the general notion, the group has identified two different levels from which the representational bias can be considered:

- the *logical* level: where aspects like the semantics of the model (imperative/declarative), the patterns to be represented, and even the ability to transform the derived models is an important issue.
- the *user* level: where quality metrics related to the user must be taken into account, e.g., truthfulness, readability, multi-perspective, are examples of this.

Both of these levels are by themselves challenging, and it is agreed that very few work has been done into guiding process discovery algorithms for satisfying them.

At the *logical* level, there are well-known examples of patterns that differentiate discovery algorithms: expressive power, concurrency, skip/duplicate activities, non-free-choiceness, hierarchy, loops or cancellation are typical examples of patterns that not all algorithms have. On the other hand, process discovery is harden by the presence of other problems like noise, incompleteness, concept drift (it seems is less stringent in practice). In addition, the granularity of event information has been identified as a problematic issue, but also the selection of parameters given a particular process discovery technique. One promising direction has been identified, which may alleviate some of the problems before: define log features that can help into making decisions and transformations into the log for improving the discovery.

At the *user* level, an issue which is important is the current situation of process discovery algorithms: do the current users know their bias? the discussion group has identified as a challenge the user explanation of each algorithm's bias. Apart from that, other factors like execution time available for discovery, expected truthfulness of the model, or desired

readability are crucial factors that the user may want to determine when using an algorithm. Again, very few techniques possess the aforementioned abilities.

General challenges have been identified, which listed below:

1. How to control the representational bias of process discovery algorithms? the conclusion of the discussion group was that it seems an important aspect, although few techniques offer it. One example is the ability of certain algorithms, like the inductive miner or the miners implementing the theory of regions, to focus the search of a model to certain quality criteria established a priori.
2. Meta-discovery: both at the logical and at the user level, it seems relevant to decide generically the representation. An example now comes handy: declarative models are known to be good in “turbulent” scenarios, while imperative models may better fit structured scenarios. One can use domain knowledge, log features or the like to decide it. The same can apply at a more concrete level, like what particular formalism may be better for the user, e.g., BPMN or Petri nets.
3. Industrial bias: the fact that industry is mostly considering standards like BPMN does not mean that process discovery algorithms should only aim at discovering these models. It is better to concentrate on the identification of patterns that may then be translated to the visual representation in terms of a particular formalism. Also, the group has identified the importance of having transformations between formalisms, even in the presence of precision losses or similar inaccuracies.

As starting point for further actions, the group has created a Dropbox folder where related papers will be collected in order to iterate over the literature and find synergies. As future work, it may be possible to trigger some collaborations in different dimensions (writing a report, joint efforts, and the like).

4.4 Discussion Session 4: Evaluation of Process Mining Algorithms: Benchmark Data Sets and Conformance Metrics

In this discussion session, we considered the maturity of the process mining community. While the research on process mining is maturing, the need arises for a clear benchmarking methodology, such that (a) researchers can *objectively compare* their results and performance against other researchers results and (b) researchers can *easily exchange* comparison results. The methodology should be language independent and community accepted.

During the discussion, we established that a model, when drawn by a process modeling specialist, is always created for a particular purpose. The purpose of a model should always be considered by process mining researchers in their evaluations, i.e. comparing process mining techniques should only be done for those techniques that serve the same purpose. In many papers today, comparisons are made without looking at the purpose of the models, thus leading to false comparisons. During our discussion, we identified several purposes, such as:

- Prediction, i.e. answering “what if?” questions,
- Happy flow discovery, i.e. visualizing the main process flow,
- Perfect representation, i.e. models that very accurately show what behavior was observed,
- Performance analysis, i.e. models that provide insights into the performance of an “as-is” situation, and
- Deviation discovery, i.e. where models explicitly show deviations from reference models, business rules, etc.

Depending on the purpose of a model, several aspects of the model may be more or less important, hence when evaluating the quality of a process model against an event log, several dimensions should be considered. During the discussion, we identified the following dimensions for which we believe language independent metrics should be developed:

- Replay fitness, i.e. the fraction of the observed behavior that fits a model,
- Precision, a measure for the amount of behavior allowed by a model, but not observed,
- Simplicity, which quantifies the understandability of the model given the behavior it expresses,
- Generalization, which quantifies to what extent the model generalizes from the observed behavior, and
- Level of decomposition, hierarchical or otherwise, quantifying how “flat” a model is.

When comparing process mining techniques, it is important to realize that optimality can often not be reached in all dimensions, i.e. models may be Pareto optimal. During the discussion, we came to the conclusion that there is currently no clear methodology for evaluating process mining techniques. Therefore, a full methodology will be proposed in a paper to be written by several participants of the Dagstuhl seminar.

To conclude, some clear points were made that should be considered already today when comparing process mining techniques to existing work:

- Compare new techniques with *all* existing techniques serving the *same purpose*,
- Compare new techniques against *many, randomly generated datasets*,
- Compare new techniques on *public, real-life datasets* available in the 3TU datacenter and
- Always *publish synthetic data* in the 3TU datacenter and preferably publish real life data too.

4.5 Discussion Session 5: Advanced Topics in Process Discovery: On-the-fly and Distributed Process Discovery

The goal of this discussion was to identify the main challenges posed by on-the-fly and distributed process discovery.

On-the-fly process discovery requires the compliance to the typical stream processing constraints: i) since it is impossible to store the complete stream, only a finite memory budget is allowed; ii) backtracking over a data stream is not feasible, so discovery algorithms are required to make only one pass over data, taking bounded time per event; iii) it is important to quickly adapt the process model to cope with evolving processes (concept drift); iv) the approach must deal with variable system conditions, such as fluctuating stream rates. Some on-the-fly discovery algorithms able to generate control-flow models and DECLARE models have already been defined. The success of these algorithms has been evaluated using traditional metrics defined for off-line process discovery. Thus it was debated how to define a proper evaluation measure for stream discovery tasks. It was agreed that the use of some of the data from the stream to evaluate fitness, precision, and other already defined measures can be considered a satisfactory solution, especially if these measures are then integrated over time. After some discussion, the addition of the social and data perspectives was recognized as an important first challenge. Considering on-the-fly both social and data perspectives is not obvious since the process model may change over time (concept drift). It was suggested that, concerning the data perspective, a possible solution could be to define a stability index over the control-flow and when the control-flow is stable, learn rules for choice points. When this problem is considered from a declarative point of view, a critical issue is whether the

concept of “activation of a constraint” is still “valid” in a stream setting. The result of the discussion on this issue has been that such concept seems to make sense as long as single events (disregarding the trace they belong to) are considered, while it seems not to be meaningful when considering an event within a single trace. Different declarative discovery algorithms can be devised according to whether event-based or trace-based focus is adopted.

Another important challenge that was discussed is how can discovery algorithms which are not based on simple statistics, as the already proposed algorithms, be extended to cope with the stream scenario constraints. The discussion suggested two kind of answers. A first suggested general approach has been to face a single constraint at time, so to evolve versions of the algorithm that eventually will be able to cope with all of them. A more specific approach could be to use a model update strategy that works only on the parts of the model that are affected by the current event (as already suggested by some authors in similar scenarios); however, it is not clear that the constraint on computational time will be satisfied. In addition, a potential problem of this latter approach is that concept drift, e.g. in seasonal processes, may not affect fitness while seriously affecting precision. A suggested way to cope with this issue is to adopt strategies to recognize which parts of the current process model is not used anymore and then remove them.

The discussion then focused on what should be visualized as output by these type of discovery algorithms. It was observed that it is not sufficient to output only the current model. It is more informative to display a model where not recently used parts are identified by a “cold” color, while most recently used parts are identified by a “hot” color. This allows for a comprehensive summary of historical behaviors. Moreover, it could be nice to generate a “movie” showing the evolution of the model in time. A grand challenge would be to mine the model evolution to extract a summary of how the model has evolved in time.

The discussion then turned on the usefulness of on-the-fly discovery algorithms in practical applications. There was a general agreement on the fact that first of all, companies are more and more adopting information systems able to produce and to process streams of data; secondly, this kind of algorithms are anytime algorithms which can be used under user defined time and storage constraints: user does not want to wait hours to discover that she/he selected the wrong data for process discovery. Thus, there is a positive side-effect in designing stream discovery algorithms, since this will allow the user to significantly shorten the exploration of event logs.

A final discussed issue concerning on-the-fly process discovery was whether GPUs can be used to speed-up computation for more demanding discovery algorithms. The discussion did lead to three outcomes. First of all, it was observed that first results on computing fitness for traces in a log are negative, mainly because it is very time consuming to transfer data from RAM to GPU global RAM (GRAM); moreover alignment is done in a sequential fashion. Evolution of GPU architectures and a smarter way to perform alignment could improve the situation in the near future. Secondly, computation of fitness could be distributed over many CPUs, while GPU computing can be used for other computations which are more suited for GPUs architecture. Finally, one solution to the above problems could be to study whether computation in discovery algorithms can be cast in a mathematical form which is amenable to fast GPU computation, such as matricial computation.

The second main argument of discussion was the possibility to distribute computation of process discovery (and conformance checking) algorithms over many CPUs. In fact, recently it was observed that Petri nets can be decomposed (under specific mild constraints) into small parts, so to allow distribution of computation. An interesting observation is that such decomposition can be inherited by any log generated by the target Petri net. This

allows to define a distributed discovery algorithm where the log is first decomposed into several small parts; these parts are then used to discover corresponding process models; the discovered process models are then glued together and eventually the resulting process model is simplified to obtain the final process model.

An important challenge is how to partition the log so to guarantee that the “right” process is discovered. It was observed that there are two possible ways to partition the log. A trivial one is to partition the log horizontally, i.e. a different subset of traces is assigned to each CPU core. An alternative way is to split it vertically, i.e. split each trace in several pieces and distribute them among the CPUs cores.

It is not clear, however, how to do the vertical partitioning in an optimal way, i.e. by reducing the computational effort while obtaining the correct model. A final raised question was whether conformance checking can take advantage by distributed computation. The answer was affirmative, Trivially, each CPU core has a copy of the model and checks one trace; results are finally aggregated.

In summary, on-the-fly and distributed process discovery constitute very useful techniques which pose several computational challenges. Promising research lines to successfully face these computational challenges, however, emerged from the discussion, and there is concrete hope that very soon new and more efficient and effective process discovery algorithms will be devised.

4.6 Discussion Session 6: Process Mining and Big Data

The idea of this session was to discuss the relation between (a) Big Data, (b) Big Data technologies, and (c) process mining. We now live in a time where the amount of data created daily goes easily beyond the processing capabilities of nowadays systems. Nevertheless the strategic importance of the knowledge hidden in such data, for effective decision making is paramount. The ability of organizations, governments but also individuals to collect information in a plethora of different systems/formats has largely overwhelmed the ability to extract useful knowledge from it⁴; not to cite the attempt to integrate such knowledge with relevant information available outside organizational boundaries. The rapidly growing data sets with event data provide opportunities and also challenges.

The session started by introducing to the audience the term Big Data as “the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization.” (Wikipedia). The group discussed about the need for a big data approach in process mining and the availability and existence of so called big data event logs. A part of the group was arguing on one hand that there is not really a urgent need for applying big data techniques to Process Mining since the research community is still focusing on solving other issues and that anyway big data sets are not easily accessible; on the other hand the fact that enterprises, governmental organizations and the likes are storing increasing amount of data and the process mining approaches and algorithms need to be adapted to this situation in order to be effective in the real world.

⁴ Gross, B. M. (1964). *The managing of organizations: the administrative struggle*. New York: Free Press of Glencoe.

The group identified the possibility for the research community to access large event logs from no profit organizations, governmental institutions, supply chain processes where data need to be shared. The first action point of the discussion was set to identify big data logs and made them publicly available for the Process Mining research community. Then the discussion moved on the three Vs (Volume, Velocity, and Variety) which define the dimension of BigData set in relation to Process Mining. We focused on measuring the event logs in relation to the three Vs: the group identified that big Volume in logs can mean a big number of process execution traces and/or big number of events per trace and/or big number of attributes per process/event. Big Velocity means in Process Mining that the logs needs to be processed before a given time, the rate between the incoming logs and the consumed ones need to be constant and this is valid also for the process mining algorithms. Finally we discussed about what Variety means in process logs; one obvious observation was on the fact that logs can have multiple formats and the systems should be able to deal with this, but a less obvious comment was on the fact that logs can have multiple points of view and by changing the identifier of the process case the process can be seen from a completely new perspective. Also the group moved on discussing if going full data makes sense: the trend is nowadays is a “throw in all” approach but this needs to be carefully done by analyzing costs versus benefits of this trend. Moreover it was pointed out that with big data in the picture it becomes paramount to help the user to “find the needle in the haystack” and so local or partial mining/visualization techniques may become necessary in the future. Then, after a short introduction to the main big data technologies the discussion focused on what technologies can be relevant to Process Mining: depending on the problem to solve Map Reduce can be used or not but it needs to be carefully planned because forcing a map reduce approach can easily degrade performances.

Map Reduce has been used in some cases for preprocessing the logs for correlation however the Map Reduce framework imposes some relevant constraints on the way the conformance checking or process discovery algorithm access the log data. The particular Data Partitioning step required for distributed process mining is the main reason why Map Reduce cannot be easily used for the generic approach in distributed process discovery and conformance checking. The group identified the fact that map reduce can be used for some simple Process Mining algorithm such as the Alpha algorithm⁵, other more complex algorithm, especially the ones sharing global states cannot be easily implemented in map reduce and therefore a shared memory approach (memcached, grid computing, GPUs) is advisable. Problems like concept drift on streams of events can be solved using a distributed stream processing approach (such as Storm). The discussion then moved on presenting a set of research works on distributed mining⁶ that can be considered the actual state of the art. Some approaches⁷ use distributed computing to speed up Process Mining algorithms such as the genetic process mining, however the log is replicated across the nodes and therefore this approach is not possible is the logs cannot be stored entirely in one machine. Therefore the group focused on the fact that the partitioning of the Logs is of extreme importance for effective distributed process mining. Moreover horizontal partitioning technique provides some additional benefits

⁵ A paper on this aspect titled “Big Data meets Process Mining. Implementation of Alpha algorithm in Map Reduce” will be published at EE track ACM-SAC 2014)

⁶ W. M. P. van der Aalst. Decomposing Petri nets for process mining: A generic approach. *Distributed and Parallel Databases* 31(4):471-507, 2013.

⁷ C. Bratosin, N. Sidorova, and W. M. P. van der Aalst. Distributed Genetic Process Mining. *IEEE World Congress on Computational Intelligence*, pp. 1951–1958. IEEE, 2010)

like data compression. Also Process Cubes⁸ can benefit from a distributed approach in order to speed up the slicing, dicing, drilling down and rolling up of process traces and distribute the mining of separate set of process traces. Finally the discussion ended by introducing the concept of open data sets representing all sort of data (weather, biological, traffic records, ...) which nowadays are publicly available; some Data Mining tools such as rapid miner started providing plug ins in order to use this type of data in the mining process, but this sort of data have never been used for process mining therefore it may be interesting to see if this data can provide benefits especially in the analytical aspect.

4.7 Discussion Session 7: Process Mining in Health-care

Health care is considered an interesting and promising domain for process mining application, due to its challenging processes, where significant impact can be made. The discussion took as a starting point two different views of a medical process:

- *Clinical view*: actions done for affecting the current physical state of patients. The emphasis of this view is on curing patients, improving their life expectancy and quality, and being able to predict outcomes of treatment. Treatment processes should comply with clinical guidelines.
- *Logistic / administrative view*: execution of the medical process using resources over time, spending and gaining money. This view emphasizes KPI and resource optimization, while meeting standards and constraints. It also addresses scheduling, costing, billing, and mitigating legal risks.

Process mining research has so far mainly addressed the logistic view. The Frequently asked questions of process mining in health care⁹ have relevance for both views, but their essence is at the logistic view. For the clinical view, data as well as control flow and data perspectives should be emphasized.

Current process mining approaches are capable of meeting most of the needs of the logistic view. Hence, this view poses an opportunity for the process mining community to make a significant impact and show good results.

The clinical view has so far received less attention. The challenges it raises are many fold. First, it requires addressing the data perspective, so in addition to considering the actions, their outcomes should also be addressed (e.g., the result of the X-ray). Second, understanding the data requires domain knowledge, hence collaboration with physicians is needed. Third, compliance should be assessed with respect to medical knowledge (clinical guidelines), whose representation requires expressiveness beyond that of business process models (e.g., temporal constraints). Finally, mining results need to be visualized in a way which captures all the relevant aspects and is meaningful to domain experts. Of current output forms, the output of declarative decision mining can be suitable, especially if transformed to natural language representation.

Nevertheless, important results can emerge from mining the clinical view. Specifically, these results can provide improved decision support for physicians. Furthermore, conformance

⁸ W. M. P. van der Aalst. Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining. Volume 159 of Lecture Notes in Business Information Processing, pp. 1–22, Springer 2013.

⁹ R. S. Mans et al. “Process mining in healthcare: Data challenges when answering frequently posed questions.” Process Support and Knowledge Representation in Health Care. Vol. 7738 of Lecture Notes in Computer Science, pp. 140–153. Springer Berlin Heidelberg, 2013.

checking is of importance since different decisions might be taken by different physicians in similar situations. Using current technology requires much data preparation, including preprocessing and cleaning (e.g., combining activities of the same type), dynamic labeling, and tagging events during execution.

Case Study on a Treatment of “Urinary Tract Infection”. Considering a given process model, two groups (for the clinical and the logistic view) discussed the questions to be addressed and the desired results of mining.

Logistic view: The analysis process should include three different phases: (1) initial information gathering from experts: the problem, a normative process model, KPIs. (2) from detailed questions, detailed analysis can be performed, and (3) the results can lead to further exploration. The specific questions would depend on the purpose of analysis and on the stakeholder it should serve (e.g., hospitals seek to maximize throughput and minimize costs). Benchmark data and relevant standards will be needed. Example questions may include: (1) what is the cost of every test, which tests and what is their order. Analysis can contribute to final decision (skip tests, reordering); (2) how long for each step; (3) are there reworks (tests that are repeated); (4) are things that are not recorded not necessarily happening; (5) are actions done in batches; (6) what is the average time between tests / scheduling constraints / how long are patients waiting; (7) are there differences in treatment path among physicians.

Clinical view: This group devoted less attention to the details of the specific case study, and discussed clinical view mining requirements in general. Tooling – representation should be dynamic and interactive, capable of handling spaghetti-like processes, and allow switching between views (e.g., showing simple flowchart and projecting other data). New views might also be needed. The role of context is crucial. Context includes patient data attributes (e.g., age), treatment history, and status of other running instances. To support this, mining can relate to feature selection for extracting context. Then all mining can relate to context: treatment pathways, treatment outcomes (with respect to context-dependent goals), forecasting and operational support (e.g., possible consequences of treatment options). Decision support should provide recommendations by a case-based system.

4.8 Discussion Session 8: Security and Privacy Issues in Large Process Data Sets

This session considered security and privacy issues from two perspectives. Firstly, how process mining techniques may help to secure systems and, secondly, the security and privacy issues that arise as a consequence of process mining.

Enterprise system security can be characterized in terms of the security *controls* that are required to mitigate the *threats* to the objectives of its business *processes*. Threats can range from failures in business processes to more infrastructure-level vulnerabilities such as those cataloged by vulnerability databases. Process Mining can help with threat identification by generating reference models of normal behavior against which anomalous behavior in logs can be detected and explained (conformance checking).

The application of Process Mining to the configuration and selection of security controls was discussed. Organizational mining can help in the discovery of RBAC configurations and it was suggested that process discovery could help in the discovery of behaviors used for the configuration of task-based security policies. While a discovered process can be used for subsequent conformance-checking, an alternative viewpoint is whether it is possible to also

use this process to generate/recommend security controls that enforce conformant behavior. For example, using a discovered document workflow to help deploy checksum-based controls in a document handling system. A related question is how Process Mining could be used to explore whether the current security controls are resilient to changes in the process, and vice-versa.

Audit procedures test the efficacy of security controls and it was suggested that Process Mining could provide a basis for a more complete check on control efficacy. Security controls and their audit procedures are not necessarily integrated into business processes. For example, a procedure that regularly searches the file system for stray plain-text credit card numbers operates independently of credit-card based transaction processes. A research challenge is how security controls, along with their audit procedures, can be correlated with discovered business processes in order to provide more threat-aware conformance checking.

Notwithstanding the conventional integrity, availability, authenticity, non-repudiation and confidentiality challenges surrounding process and log data, Process Mining introduces particular assurance issues. Log data can come from different sources with varying degrees of assurance and trustworthiness. One question is how these relationships might be securely managed and how they might be reflected, not just in the original log-data, but also surfaced into any discovered process. Would such a scheme require a single security authority with jurisdiction over all log data and sources, or can a more decentralized approach be taken? The latter may be useful if organizations share log data. For example, organizations merging or aligning their interdependent processes in a supply-chain. For these federated logs, what are their security and privacy requirements and how might they be implemented?

Different users may have different views on different process logs and process mining should preserve these view restrictions. A challenge is the extent to which Process Mining can be carried out on views alone rather than on the full log-data. An advantage of the former is that any (security) failure in mining does not expose data outside the view, while the latter provides more precision but requires assurance in the process mining software. A further challenge is how log data can be reliably de-identified/anonymized. Differential-privacy based techniques may be useful in implementing privacy aware views: a discovered process should not reveal data previously de-identified in the log.

Lastly, a recent Semantic study on the cost of data breaches identified human factors and business process failures as a significant contributory factor. Given that Process Mining helps provide deeper understanding and control of business processes it would be worthwhile investigating its application to identifying process weaknesses that may lead to data breach.

4.9 Discussion Session 9: Conformance Checking for Security, Compliance and Auditing

A large part of the discussion was concerned with coming to an agreement of the terminology. The discussion was very active but, unfortunately, it has been impossible to agree on the terminology. In particular, it was clear to identify two different schools of thought:

- A first school differentiated between a-priori and a-posteriori verification. Compliance checking is concerned with verify the TO-BE model against norms, regulations, security constraints. Therefore, compliance checking is strictly related to analysis and verification of an executable model, which should adhere to constraints imposed by laws and regulations. In many settings, it is not strictly enforced that the actual execution follows this model. Here, conformance checking comes to play. Conformance checking is about verifying

whether the actual executions, as recorded in the event log, follow the same constraints. In this case, there is not one single model but, rather, multiple models that are built ad-hoc for the sole purpose of rules' checking.

- A second school sees conformance checking as a mean to check compliance. Conformance Checking takes a regulatory model and an event log and highlights the non-conformances. In this sense, there is some common point with the view at point 1. The serious difference is with respect compliance checking, which is still about verifying the behavior observed in the event log. Compliance checking is concerned with a number of norms, regulations that are converted in a number of regulatory models. Using the conformance-checking means, each of the regulatory models is verified against the event log at disposal. In this second school, auditing is an umbrella under which compliance checking is placed, along with “static” conformance verification. The latter refers to an executable model that is wanted to check the adherence with laws and regulations.

During the discussion, many issues raised up about the languages. In the last decade, several languages have been standardized. The discussion's participants agreed on the fact that there are several standards but not one largely-recognized standard.

Some languages are characterized by a precise syntax and semantics but a bit complex to use for average process analysts. Some are more user-friendly but at the cost of less accuracy in the semantics. This raised another important point: the usability of the languages. The language complexity can be problematic if process analysts do not have a sufficient background in Mathematics. Therefore, the syntax should be kept simple and intuitive for moderately-skilled process analysts but, at the same time, should have a precise and well-defined semantics to not lead to multiple interpretations. Furthermore, decidability and complexity of the algorithms that are used for conformance/compliance checking put a bound on the expressiveness of the languages used to define the conformance/compliance rules.

Once a language and compliance/conformance algorithms are chosen, it is possible to check for compliance and conformance. Of course, a pure YES/NO answer is not enough, i.e. the executions of single process instances are compliant/conformant with rules and regulations. The discussion participants acknowledge the important of pinpointing deviations and their root-causes. Moreover, in many settings, deviations do not occur in isolation: a deviation may cause a cascading effect, which can lead to further deviations. Hence, it is important to relate to each other when a relation may exist.

A few other important thoughts were discussed during the session. Firstly, in continuous auditing, it is important to be equipped with decision support that guides participants to not violate compliance constraints at run time. Using event logs, machine-learning techniques can be used to discover the common patterns that lead to problems and, hence, the decision support system can suggest execution paths that avoid them. Problems of applying runtime reasoning, e.g. Supervisory Control Theory, are rarely applicable in this context as the problem is inherently hard and, often, become undecidable. Secondly, in many settings, primarily in Security, the conformance/compliance of the execution of process instances cannot be checked in isolations. The compliance of an instance may depend on other instances of the same process or, even, on instances of different processes. This also highlights the importance of contextual information, which should be incorporated into the analysis.

As a conclusion, every discussion participant agreed on the fact that checking compliance/conformance is an important topic in the field of business process management and auditing. The main issue seems that, unfortunately, there is not a large consensus on what checking compliance/conformance actually means. An effort should be made in order to make sure that there exists no different wording for the same concepts. The discussion was

very animated and live. This certifies that checking compliance/conformance is certainly a topic that will attract much attention in the future. As a matter of fact, many of the existing techniques are still in a development stage and there is a lot of room for future improvements.

4.10 Discussion Session 10: How to Sell Process Mining?

Goal: Since everything in selling is about understanding the customer, the purpose of the discussion was to first gain a more nuanced view about the different types of possible target customers for process mining.

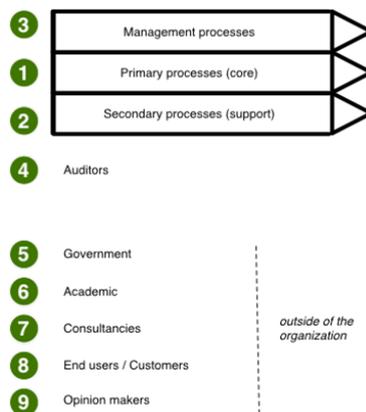
Approach: We collected a broad range of target customers from the group and then discussed three profiles in more detail. Afterwards, the results were put together by Frank van Geffen and Anne Rozinat.

Results: An operational manager of a business process is confronted with different (sometimes conflicting) goals, as depicted in Fig. 5.



■ **Figure 5** Conflicting goals.

We used the value chain-model of Porter to categorize the various business functions we had collected and added the other categories outside of the organization. The value-chain model is depicted in Fig. 6.



■ **Figure 6** Porter's value-chain model.

The following target customers for process mining were mentioned in these categories:

1. Operational manager, Sales department, Customer satisfaction representative, Process managers, Requirements engineer, Operational people, Process manager / Department head, Knowledge worker, Software product development, Supply management

2. Business consultants, CIO, Quality and process improvement department, IT managers, Process management department, Business analyst, Development and analyst people, Knowledge worker
3. Higher managers, Business controllers, CEO
4. Financial auditors, operational auditors, IT auditors, Crisis and fraud people
5. European commission
6. Other scientists
7. Consultancy firms
8. End user
9. Analyst firms, such as Gartner and Forrester.

We used the following three profiles to paint an as detailed picture of the respective customer as possible:

- A Business analyst at a hospital
- An End user (customer) of a service organization like the Rabobank
- An Auditor at a manufacturing company like Boeing

For each of these profiles we then tried to answer these questions: “How do they spend their day?”, “Which processes are in their sphere?” “What are their challenges?”, and “How can process mining help?”

4.11 Discussion Session 11: What is the Ideal Tool for an Expert User?

The idea of this session was to discuss the requirements and ideas for realizing process mining tools aiming at experts. The discussion focusses on the following three goals: identification of expert user types, identification of functional and non-functional requirements and suggestions how the academic community could support the development of expert tools.

Since “expert user” is a rather vague and general term, the discussion started with the identification of different expert user types. Two dimensions were proposed to identify different experts, i.e. the problem and user dimension. The problem dimension divides process mining problems along a range from well structured, standardized and repeating problems to unique, ill-structured and more generic problems. As for the user dimension, a distinction should be made between developers of process mining algorithms and actual users of process mining algorithms. Based on these two dimensions, three expert users could be identified.

The *Algorithm Developer* creates new process mining algorithms for generic process mining problems. Typically, this type of expert user can be found in academia. The *Data Scientist* can be found in a business environment and solves ill-structured process-related business questions by means of data and the use of existing process mining algorithms. The *Tool integrator* is also a business user, but in contrast to the *Data Scientist*, deals with structured and reoccurring process mining problems. They typically develop tool chains of process algorithm tools, business intelligence tools and enterprise systems to generate management dashboards.

Next, the discussion continued with the identification of non-functional requirements, which are the tool-usage specific properties. The *Algorithm Developer* requires a tool that allows reuse of existing algorithms, the ability to modify existing algorithms, proper documentation with instructions on how to develop with the expert tool, a system which

encourages or even enforces proper documentation of newly developed algorithms and a comprehensive overview of all available algorithms.

The *Data Scientist* requires a tool that provides scripting functionality and an easy integration with third party applications. The tool should allow the expert user to easily import and export data from and to a wide range of formats and should provide a flexible environment to manipulate data. Finally, the tool should offer algorithms that are robust and fast, well documented from a user's perspective and which can be tuned by means of parameters.

The Tool Integrator requires a tool that provides solid interconnectivity opportunities with various enterprise systems. If the tool aims to generate the dashboard directly, it should provide a flexible and customizable user interface. Furthermore, it would be a great asset if the tool already provided and supported predefined workflows for standard process mining problems.

Next, the discussion focused on the functional requirements which refers to currently unavailable algorithms which will become increasingly important in the near future. The discussion group identified the need for process mining algorithms which allows for: distributed process mining, data stream process mining, predictive process mining, multi-perspective process mining, direct comparison of processes at the process model level, process simulation, interactive visualization.

Finally, the meeting agreed on the importance that advances in process mining research gets integrated in expert tools and the need for the scientific community to contribute to the development of expert user tools. However, the concern was raised that currently there are little to no incentives for academia to actively contribute to the development of expert tools. For example, the current log loader of ProM has problems with loading big data, which hinders the further development of process mining algorithms for big data. While the community would clearly benefit from the development of a new log loader, it is in no individual's academic interest to spend a lot of time on this. To solve this catch-22, it was suggested that the research community agrees to assign budget in future (European) research projects to the development of much required, but academically non-interesting features and components of tools for algorithm developers.

4.12 Discussion Session 12: What is the Ideal Tool for a Casual Business User?

The idea of this session was to discuss the requirements and key success factors of a process mining tool tailored to business users. To this end, the following topics were discussed:

- What are the essential functionalities of a process mining tool for the casual business user?
- What is the functionality in existing tools that is most useful?
- What is missing?
- Should tools provide operational support (on-the-fly discovery, prediction, checking, recommendations)?
- How to visualize results?

In essence, the discussion could be capture as follows: vendors sell *features* while customers see the *benefits*. Given this line of discussion, Table 1 depicts the relationship – worked out during the discussion – between different types of uses and the corresponding benefits. Table 2 establishes the relationship between the benefits and the expected functionality.

■ **Table 1** Relationship between type of user and expected benefits.

User	Benefit
Process Owner	Verify if employees meet the rules she has in mind
Manager of document flow department.	Show if the operations meet regulation and SLA: case must be closed in 30 days
Scientist / researcher in hospital	Interested in finding compliance to medical guidelines
M&A director/team, HR director	Want to understand processes in order to integrate companies.
Auditor	Auditor wants to find unknown patters. Auditor knows what rules should be followed and wants to test if those rules are followed.
Operational manager of customer service process	Be proactive and improve processes
Process Innovation department/team reports to CEO	Optimize the performance of whole corporation
Security Auditor	Find evidence of fraud in transaction log
Process Owner / director of service delivery	Improve operational performance

■ **Table 2** Relationship between benefit and expected functionality.

User + Benefit	Functionality
Process Owner: Verify if employees meet the rules she had in mind	Compliance/conformance checking: list of cases that violate the tules
M&A director/team, HR director: Want to understand processes in order to integrate companies.	Use process mining discovery, show workflow, organization perspective, data flow diagrams.
Operational manager of customer service process: Be proactive and improve processes	1. deeper understanding of process...tool should give metrics , New measure values should come on weekly bases automatically. Consultant does the initial work and then business user can get what she wants by herself. 2. Operative management wants to get graphs and dashboards . And then predictive analysis ...imagine a metric in a process...if we keep going like this then I will meet my objectives ...
Auditor:	Model checking feature like 4-eye principle. Feed in business requirements and then show cases that meet the requirements and those which do not.
Security Auditor: Finding evidence of fraud in transaction log	They want to have simulation/what-if that comes up with possible changes in the processes and suggests the best mitigation action.

The discussion session ended up the following insights: An ideal tool. . .

- . . . gets source data easily and with high quality. If the data does not come then there is no continuous benefits.
- . . . actively supports the user and shows only relevant options and functions.
- . . . supports interactive navigation during discovery phase.
- . . . is configured by the consultant and used by the business user.
- . . . has certain set of flexibility for Business End Users. They do not want to call always consultant to make changes.

Currently the users are typically the early adopters who are willing to use a complex tool. Typical business managers do not have time to play with the tool.

Participants

- Rafael Accorsi
Universität Freiburg, DE
- Antonio Caforio
University of Salento, IT
- Josep Carmona
UPC – Barcelona, ES
- Paolo Ceravolo
University of Milan, IT
- Jan Claes
Ghent University, BE
- Jonathan Cook
New Mexico State University, US
- Ernesto Damiani
Università degli Studi di Milano –
Crema, IT
- Massimiliano de Leoni
University of Padova, IT & TU
Eindhoven, NL
- Benoit Depaire
Hasselt Univ. – Diepenbeek, BE
- Simon N. Foley
University College Cork, IE
- Luciano Garcia-Banuelos
University of Tartu, EE
- Christian Günther
Fluxicon Process Lab., NL
- Anna A. Kalenkova
NRU Higher School of
Economics – Moscow, RU
- Akhil Kumar
Pennsylvania State Univ., US
- Geetika T. Lakshmanan
IBM TJ Watson Res. Center –
Cambridge, US
- Teemu Lehto
QPR Software – Helsinki, FI
- Marcello Leida
Khalifa Univ. – Abu Dhabi, AE
- Maria Leitner
Universität Wien, AT
- Fabrizio Maria Maggi
University of Tartu, EE
- Ronny S. Mans
TU Eindhoven, NL
- Alexey A. Mitsyuk
NRU Higher School of
Economics – Moscow, RU
- Jorge Munoz-Gama
UPC – Barcelona, ES
- Zbigniew Paszkiewicz
Poznan Univ. of Economics, PL
- Alessandro Provetti
Università di Messina, IT
- Hajo A. Reijers
TU Eindhoven, NL
- Joel Tiago Ribeiro
UPC – Barcelona, ES
- Roland Rieke
Fraunhofer SIT – Darmstadt, DE
- Stefanie Rinderle-Ma
Universität Wien, AT
- Anne Rozinat
Fluxicon Process Lab., NL
- Ricardo Seguel
Excellentia BPM, CL
- Marcos Sepulveda Fernandez
Pontificia Universidad Catolica
de Chile, CL
- Sergey A. Shershakov
NRU Higher School of
Economics – Moscow, RU
- Pnina Soffer
University of Haifa, IL
- Minseok Song
UNIST, KR
- Alessandro Sperduti
University of Padova, IT
- Thomas Stocker
Universität Freiburg, DE
- Wil van der Aalst
TU Eindhoven, NL
- Boudewijn van Dongen
TU Eindhoven, NL
- Frank van Geffen
Rabobank – Utrecht, NL
- Eric Verbeek
TU Eindhoven, NL
- Thomas Vogelgesang
Universität Oldenburg, DE
- Jianmin Wang
Tsinghua University Beijing, CN
- Barbara Weber
Universität Innsbruck, AT
- Ton Weijters
TU Eindhoven, NL



Forensic Computing

Edited by

Felix C. Freiling¹, Gerrit Hornung², and Radim Polcák³

1 Friedrich-Alexander-Universität Erlangen-Nürnberg, DE,

felix.freiling@cs.fau.de

2 Universität Passau, DE, gerrit.hornung@uni-passau.de

3 Masaryk University, CZ

Abstract

Forensic computing (sometimes also called *digital forensics*, *computer forensics* or *IT forensics*) is a branch of forensic science pertaining to digital evidence, i.e., any legal evidence that is processed by digital computer systems or stored on digital storage media. Forensic computing is a new discipline evolving within the intersection of several established research areas such as computer science, computer engineering and law.

Forensic computing is rapidly gaining importance since the amount of crime involving digital systems is steadily increasing. Furthermore, the area is still underdeveloped and poses many technical and legal challenges.

This Dagstuhl seminar brought together researchers and practitioners from computer science and law covering the diverse areas of forensic computing. The goal of the seminar was to further establish forensic computing as a scientific research discipline, to identify the strengths and weaknesses of the research field, and to discuss the foundations of its methodology.

The seminar was jointly organized by Prof. Dr. Felix Freiling (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany), Prof. Dr. Radim Polcák (Masaryk University, Czech Republic), Prof. Dr. Gerrit Hornung (Universität Passau, Germany). It was attended by 22 participants and its structure was based on experiences from a similar seminar in 2011 (Dagstuhl Seminar 11401).

Seminar 24.–29. November, 2013 – www.dagstuhl.de/13482

1998 ACM Subject Classification D.4.6 Security and Protection, K.4.2 Social Issues, K.5.2 Governmental Issues

Keywords and phrases practical experience in forensics and law; applied forensic computing

Digital Object Identifier 10.4230/DagRep.3.11.193

Edited in cooperation with Michael Gruhn

1 Executive Summary

Felix C. Freiling

Radim Polcák

Gerrit Hornung

License  Creative Commons BY 3.0 Unported license
© Felix C. Freiling, Radim Polcák and Gerrit Hornung

After a brief introduction by the organizers, the seminar started off with a sequence of 3 slide/5 minute talks by all participants stating their research interests, their background and their expectations towards the seminar. In the afternoon, three motivation talks by Felix Freiling (“What is forensic computing?”), Gerrit Hornung (“The fundamental rights



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Forensic Computing, *Dagstuhl Reports*, Vol. 3, Issue 11, pp. 193–208

Editors: Felix C. Freiling, Gerrit Hornung, and Radim Polcák



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

dimension of individual and mass surveillance”) and Radim Polčák (“Experiences from drafting the cybersecurity act in CZ”) paved the way for a common understanding of the open questions in the area and the relation of forensic computing to computer security law.

The rest of the afternoon questions and expectations were collected and grouped using moderation cards. The result was a spectrum of five areas of interest that we termed as follows:

1. technical possibilities for evidence collection
2. digital evidence: admissibility, spoofing, integrity protection
3. open source intelligence
4. investigations vs. privacy
5. offensive countermeasures
6. transborder/cloud evidence collection

For immediate discussion on Tuesday the participants voted for their favorite topics. As a result, three discussion groups were formed for the next day: digital evidence (topic 2), investigations vs. privacy (topic 4) and offensive countermeasures (topic 5). Topic 1 was to be handled by an overview talk by Andreas Dewald on the following day.

Tuesday morning started with a talk by Andreas Dewald on technically unavoidable evidence and was followed by a multimedia presentation about cold boot and hot re-plug attacks. After this technical introduction work in the discussion groups took place until the afternoon, when the collected results of the discussion groups were presented in a plenary session. As a highlight, the group on offensive countermeasures presented a taxonomy of 5 categories of offensive countermeasures that were specific enough for both law and computer science to investigate. The results of all discussion groups are summarized later in this report.

Wednesday morning commenced with a talk about the work of Interpol by Jan Ellermann (“Data protection as an asset in Europol’s fight against cybercrime”). It was followed by a presentation of current research by Dominik Herrmann about the usage of fingerprinting in network forensics (“Fingerprinting Techniques for Network Forensics”). The round of talks was concluded by an introduction to the law of evidence in criminal procedural law by Tobias Singelstein (“Basics zum Beweisrecht im Strafverfahren”).

The afternoon was spent on a pleasant hike to a nearby village where the Dagstuhl office had organized delicious traditional coffee and cake. On the way back to Schloss Dagstuhl a group of adventurers again, as in 2011, separated from the main party to explore the woods around Wadern. However, unlike 2011, they managed to return to Dagstuhl in time without major difficulties.

Thursday was started with a talk by Dennis Heinson on investigations in enterprises (“Internal Investigations, IT Forensics and Law”). Afterwards two new discussion groups were formed, partly based on the areas of interest collected on Monday, and commenced discussing the topics of (1) internal investigations and (2) transborder/cloud issues. In the afternoon, the results of these groups were collected in a plenary session during which especially the transborder issues caused a heated and insightful discussion.

Friday morning hosted a series of three talks from computer science, law and practice by Christian Hawellek (on techniques for modeling surveillance), Stefan Kiltz (“Forensically Sound Data for Digitised Forensics on the Example of Locksmith Forensics”) and – last but not least – Erich Schweighofer (“Surveillance of US-surveillance”).

Conclusion

In summary, the participants (and the organizers) enjoyed the week in Dagstuhl. In particular, the chance to get to know many new people from both the technical and the legal side of forensic computing was appreciated. From the viewpoint of the organizers, several points appear worth mentioning which we wish to document here.

First of all, it became clear to all participants that forensic computing is still in the process of maturing. The legal regulations as well as the technical instruments used in forensic computing are evolving quickly and it needs a joint effort by both communities to make progress. In our opinion, the seminar was much better than the preceding seminar in 2011, mainly because the lawyers were more interested in technical details and the technical people presented their “special secret instruments” in an understandable way. The seminar showed that fruitful discussions between both sides are possible, that lawyers can be cool as well and that there exist at least some lawyers with advanced technical understanding. For the technical people it was insightful to get a basic feeling on how the interpretation of law works and to see that there are quite a lot of gray legal areas. After all, forensic expertise is just one bit of evidence in court, and it may not be the most important one. And there are actually many, many data protection problems out there that will need to be handled within the field of forensic computing.

Overall, it was again a challenge to gather interested people in Dagstuhl. Dagstuhl seminars are well-known in computer science, but not in law, and it is well-known that practitioners, which are common in forensic computing (prosecutors, defenders, police, expert witnesses), with their tight time schedules can hardly afford to come to Dagstuhl for an entire week, especially from overseas. This is a problem which will remain and explains why – again – the seminar was dominated by German speaking participants.

The topic of forensic computing, however, is also gaining importance in the academic community, and at Dagstuhl: In February 2014, a seminar on “Digital Evidence and Forensic Readiness” (Dagstuhl Seminar 14092) will take place, opening the possibility for several of the participants to meet and discuss again, albeit with a slightly sharpened focus. In case another general seminar like this would take place, the topic of mutual understanding can be placed into focus even stronger. This could be achieved by distributing introductory papers from “the other side” in advance or by giving introductory tutorials in forensic techniques at the seminar. In the end, the seminar left us with more open questions than we had at the beginning. But at least this was to be expected.

2 Table of Contents

Executive Summary

<i>Felix C. Freiling, Radim Polcák and Gerrit Hornung</i>	193
---------------------------------------------------------------------	-----

Overview of Talks

Technically Unavoidable Evidence <i>Andreas Dewald</i>	197
Data protection as an asset in Europol's fight against cybercrime <i>Jan Ellermann</i>	197
What is Forensic Computing? <i>Felix C. Freiling</i>	198
Internal Investigations, IT Forensics and Law <i>Dennis Heinson</i>	198
Fingerprinting Techniques for Network Forensics <i>Dominik Herrmann</i>	199
The fundamental rights dimension of individual and mass surveillance <i>Gerrit Hornung</i>	199
Forensically Sound Data for Digitised Forensics on the Example of Locksmith Forensics <i>Stefan Kiltz</i>	199
Experiences from drafting the cybersecurity act in CZ <i>Radim Polcák</i>	200
Basics zum Beweisrecht im Strafverfahren <i>Tobias Singelstein</i>	201
Surveillance of US-surveillance <i>Erich Schweighofer</i>	201

Discussion Groups

Digital Evidence <i>Dominik Herrmann</i>	202
Cross-border cloud investigations <i>York Yannikos</i>	202
Forensics vs. Privacy <i>Felix Freiling</i>	204
Offensive Countermeasures <i>Thomas Schreck, Michael Gruhn</i>	204
Internal investigations <i>Christian Hawellek, Michael Gruhn</i>	206

Participants	208
-------------------------------	-----

3 Overview of Talks

3.1 Technically Unavoidable Evidence

Andreas Dewald (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Andreas Dewald

A common question in forensic computing is the following: Looking at a seized hard drive, which actions have been performed by the user? To answer this question, the forensic expert needs to analyze digital evidence and finally presented it in court. But which evidence is good evidence? How meaningful is it? How convincing? The probably most important distinction regarding this questions is the notion of technically avoidable and unavoidable evidence. Technically avoidable evidence is data that is generated for its own sake and can easily be avoided, such as document files, for example. Technically unavoidable evidence in contrast is data that is unintentionally (and most times unknowingly) generated by the system and cannot be configured away – at least not by a “normal” user. Even though this definition strongly depends on the perception of what the user is able to do, technically unavoidable evidence has a high probative value in general. As an example of such evidence, we present the concept of application fingerprinting based on filesystem timestamps: We found that on any action performed by a user (like sending an email or browsing a website) timestamps of various files are changed in a characteristic way. Those “fingerprints” in timestamp data are a good example for technically unavoidable evidence with high probative value. In this case, the extraction of the evidence even can be highly automated to support the investigation with a quick overview of known actions that happened lately.

3.2 Data protection as an asset in Europol’s fight against cybercrime

Jan Ellermann (Europol, NL)

License © Creative Commons BY 3.0 Unported license
© Jan Ellermann

Joint work of Ellermann, Jan; Drewer, Daniel

Main reference D. Drewer, J. Ellermann, “Europol’s data protection framework as an asset in the fight against cybercrime,” ERA Forum, 13(3):381–395, 2012.

URL <http://dx.doi.org/10.1007/s12027-012-0268-6>

The European Union has launched its own European Cybercrime Centre (EC3) at the beginning of 2013. A related feasibility study carried out for the European Commission reveals that next to operational considerations strong data protection safeguards constitute one of the main factors for having the centre hosted at the European Police Office (Europol). Data protection and the fight against cybercrime do certainly not constitute a contradiction. On the contrary, due protection of information relating to identified or identifiable natural persons is a prerequisite to prevent identity theft and other forms of cybercrime.

The talk I have given has illustrated the solid data protection regime at Europol. Prominent features in this regard are independent data protection supervision, Europol’s secure information exchange capabilities, data protection compliant outreach to the private sector and – most importantly – clearly defined purpose specifications for processing operations upon personal data in Europol’s databases.

The aims of preventing and combating cybercrime are balanced against the goal of safeguarding the freedom of individuals. In fact, they go hand in hand: at Europol, it is

recognised that the data protection rules in place are essential for the success of operations. High data protection standards lead to high quality of data which itself is a precondition for high quality crime analysis.

3.3 What is Forensic Computing?

Felix C. Freiling (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 3.0 Unported license
© Felix C. Freiling

We discuss the different notions of digital forensics, computer forensics and forensic computing in context and try to approach a useful meaning of the term “forensic computing”. We argue that any notion of forensic computing should be defined in comparison to fundamental theories in traditional forensic science. We propose a separation of what is currently demanded of practitioners in digital forensics into a rigorous scientific part on the one hand, and a more general methodology of searching and seizing digital evidence and conducting digital investigations on the other. We thereby mark out the route for computer forensics to turn into a true forensic science.

3.4 Internal Investigations, IT Forensics and Law

Dennis Heinson (Hamburg, DE)

License  Creative Commons BY 3.0 Unported license
© Dennis Heinson

Internal investigations in enterprises are, alongside criminal investigations conducted by state authorities, the main field in which IT forensics get performed. The legal regime that governs such investigations is substantially different to the law of criminal procedure which applies to criminal prosecution. While issues such as admissibility and relevance of evidence are largely in accord between both procedural regimes, the rules that govern the collection of evidence differ fundamentally.

For in internal investigations the private entities that collect evidence are not bound by the strict rules of the code of criminal procedure. Instead, mainly data protection laws may constrain the permissible means, scope and depth of an examination. In it, as a general principle, an investigation must be justifiable both with regards to the suspect as well as any third persons whose data gets processed during the course of the investigation. The applicable provisions contain no clear-cut criteria as to the “if’s” and “how’s” of an investigation, but instead mark the outer boundaries of what is allowed. Generally, only suspicion based on facts that an employee has committed a crime may trigger an investigation.

3.5 Fingerprinting Techniques for Network Forensics

Dominik Herrmann (Universität Hamburg, DE)

License © Creative Commons BY 3.0 Unported license
© Dominik Herrmann

Joint work of Herrmann, Dominik; Fuchs, Karl-Peter; Federrath, Hannes

Main reference D. Herrmann, K.-P. Fuchs, H. Federrath, “Fingerprinting Techniques for Target-oriented Investigations in Network Forensics,” to appear in Proc. of GI SICHERHEIT 2014, März 2014.

Fingerprinting techniques are receiving widespread attention in the field of information security. This talk shows why they may be of specific interest for the field of network forensics. Firstly, fingerprinting techniques can be used to infer the activities of suspects, even when communication is encrypted. Secondly, they can be used to associate criminal activity with a suspect, even in the absence of explicit identifiers. In order to illustrate the utility of fingerprinting techniques three case studies are introduced. For each case study the applicability of existing as well as new fingerprinting techniques, which are based on DNS queries, is reviewed. Finally, some arguments are provided in order to start a discussion about the opportunities and risks that may result from using evidence gained by fingerprinting techniques in criminal investigations.

3.6 The fundamental rights dimension of individual and mass surveillance

Gerrit Hornung (Universität Passau, DE)

License © Creative Commons BY 3.0 Unported license
© Gerrit Hornung

Forensic computing regularly interferes with fundamental rights, such as informational self-determination, confidentiality and integrity of IT systems, fair trial etc. These rights form different layers with different supervisory bodies, leading to complex material and procedural situations. This becomes even more complex when compared to the US situation, when the FISA court recently denied that so-called meta data enjoys the protection of the fourth amendment. We discussed the limitations of judicial review in the area of secret agency surveillance post-Snowden, as well as possible topics to discuss in the seminar deriving from the fundamental rights dimension.

3.7 Forensically Sound Data for Digitised Forensics on the Example of Locksmith Forensics

Stefan Kiltz (Universität Magdeburg, DE)

License © Creative Commons BY 3.0 Unported license
© Stefan Kiltz

Joint work of Stefan Kiltz, Eric Clausing, Jana Dittmann, Claus Vielhauer

In a lot of disciplines in crime scene forensics (e.g. ballistics, dactyloscopy, forensic locksmithing), research is conducted to introduce IT systems to support the forensic experts off strenuous, repetitive and error-prone task (termed as digitised forensics). Instead of analysing physical objects, digital representations of some aspects of the physical world are investigated after an analogue to digital conversion. Often this involves pattern recognition

as a means to basically enhance the contrast between a latent trace and the surrounding environment, which could employ machine learning and statistics based approaches.

Digitised forensics both brings new opportunities (e.g. when using contactless sensory little or no influence of the investigation process on the recovered artefacts) but also brings on new challenges (e.g. if latent traces are rendered visible only in the digital domain, how can the process still be comprehensible also for technical laymen such as the judge). As a direct consequence of the latter, two chains of custody need to be maintained, the conventional chain of custody for physical objects but also a new chain of custody for digital objects. As with the conventional, also the chain of custody for digital objects needs to adhere at least to the security aspects of integrity and authenticity.

To model the process (a model as a means to ensure comprehensibility), existing models for the forensic process for IT forensics can be successfully adopted. It has been researched, that the splitting into several investigation steps and their order of execution is vital to achieve comprehensibility and thus transparency by allowing to annotate the different forensic data types with the investigation process of the input data and with the investigation process of the output data. This allows to identify dependencies, which can be vital e.g. if some piece of data is identified to be bad or erroneous late in the investigation process. In such cases all other data and the conclusions drawn from its interpretation needs to be identified and marked as also inaccurate and, if possible, re-investigated.

This talk is based on the following two articles:

- Stefan Kiltz, Jana Dittmann, Claus Vielhauer: “Beweissichere Daten in der digitalisierten Forensik”, In D-A-CH Security 2012: Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven. Konstanz, Germany. 25.–26.09.2012, IT Verlag Sauerlach, ISBN 978-3-00-039221-4, pp. 288–300, 2012.
- Stefan Kiltz, Eric Clausing, Jana Dittmann, Claus Vielhauer: “Ein Vorgehensmodell für die digitale Schlossforensik”, In D-A-CH Security 2013: Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven. Nürnberg, Germany, 17–18.09.2013, syssec, pp. 367-379, ISBN 978-3-00-042097-9, 2013.

3.8 Experiences from drafting the cybersecurity act in CZ

Radim Polcák (Masaryk University, CZ)

License  Creative Commons BY 3.0 Unported license
© Radim Polcák

When cybersecurity becomes a legislative issue, there is a need to tackle it not primarily as technical but rather a social phenomenon. In that sense, the duties laid down by qualitatively new sort of legislative provisions have to be grounded on proper material legitimacy. In particular, there is no such thing as “security,” whereas there is always a need to ask as to “what” is to be secured. In the case of cybersecurity, the primary concern is to create an environment for individuals to exercise their fundamental information rights, most of all the right for information self-determination (this complex concept includes passive protective rights towards privacy and personal data as well as active rights to have an opportunity to use services of information society in order to engage in social interaction) and freedom of speech.

In the Czech Republic, as well as in the case of other post-communist countries, there had to be especially tackled issues related to content-oriented regulation and institutional backing. Experience with communist censorship as well as with the situation in which the

state is understood as an enemy of a person, enhances social and political sensitivity of the regulatory task. Consequently, the Bill, apart from being based on the idea of protection of individual information rights, was built upon the principle of technology neutrality as well as on institutional distinction between security and law enforcement. Consequently, the agenda was entrusted into the hands of the National Security Authority, whereas possible use of information gathered in the course of security operations are to be made available to other branches of state executive upon standard principles of transparent inter-institutional cooperation.

The paper, apart from explaining the aforementioned fundamental regulatory concerns, discussed also particular regulatory features of the legislative draft – namely the system of reporting and processing of data on security incidents, cooperation between national and governmental response teams as well as possible consequences of drafted EU common regulatory framework.

3.9 Basics zum Beweisrecht im Strafverfahren

Tobias Singelstein (Freie Universität Berlin, DE)

License  Creative Commons BY 3.0 Unported license
© Tobias Singelstein

Das Strafverfahren gliedert sich in drei aufeinanderfolgende Phasen. Es beginnt mit dem in der Praxis äußerst wichtigen Ermittlungsverfahren, das von der Staatsanwaltschaft geleitet wird. Hieran schließen sich das Zwischen- sowie dann das Hauptverfahren an, wo die Verfahrensherrschaft beim Gericht liegt. Für den Beginn eines Ermittlungsverfahrens ist ein Anfangsverdacht erforderlich, d.h. konkrete tatsächliche Anhaltspunkte, dass eine Straftat begangen worden ist. Will die Staatsanwaltschaft am Ende des Ermittlungsverfahrens Anklage erheben, so benötigte sie hierfür einen hinreichenden Tatverdacht, also die Wahrscheinlichkeit einer Verurteilung im Hauptverfahren. Das Gericht schließlich benötigt für sein Urteil am Ende des Hauptverfahrens die richterliche Überzeugung von der Schuld des Angeklagten. All dies – Anfangsverdacht, hinreichender Tatverdacht, richterliche Überzeugung – wird aufgrund von Beweismitteln beurteilt, die vor allem im Rahmen des Ermittlungsverfahrens gesammelt werden.

Dabei sind die Strafverfolgungsbehörden einerseits verpflichtet, den Sachverhalt bestmöglich aufzuklären. Andererseits unterliegen sie rechtlichen Grenzen. Ermittlungsmaßnahmen zur Beschaffung von Beweismitteln stellen in aller Regel Grundrechtseingriffe dar. Solche Eingriffe sind nur aufgrund einer hinreichend bestimmten gesetzlichen Grundlage zulässig. Fehlt es an einer gesetzlichen Grundlage oder sind die Grenzen dieser Grundlage im konkreten Fall überschritten worden, so ist die Beweiserhebung rechtswidrig und die erlangten Beweise dürfen unter Umständen im Prozess nicht verwendet werden. Dies ist der Fall wenn ein so genanntes Verwertungsverbot vorliegt.

3.10 Surveillance of US-surveillance

Erich Schweighofer (Universität Wien, AT)

License  Creative Commons BY 3.0 Unported license
© Erich Schweighofer

This talk presented a discussion outline of a legal assessment of US-surveillance in Austria.

4 Discussion Groups

4.1 Digital Evidence

Dominik Herrmann

License  Creative Commons BY 3.0 Unported license
© Dominik Herrmann

The discussion group on the topic of “Digital Evidence” consisted of participants with a background in law as well as participants with a background in information security and digital forensics.

The group looked at particular issues related with the collection and interpretation of digital evidence during criminal investigations. The discussion started out with four main questions: What sorts of evidence can be collected? Is the collected evidence good enough to substantiate specific claims? How can digital evidence be collected, i. e., what are the requirements for evidence collection procedures? What types of digital evidence are law enforcement agencies allowed to collect?

Participants felt that especially the “problem of probabilities”, i. e., high accuracy values given for newly proposed forensic techniques in scientific papers are an issue. High accuracy values convey a false sense of confidence, although the techniques have never been put to the test in a practical environment. As a means to address this issue, participants suggested to work on the standardization of collection and analysis procedures in practical settings.

The discussion raised two questions: Firstly, there was a debate whether digital evidence can be fabricated or faked more easily due to the fact that information does not degrade when it is copied. On the other hand, it may be more difficult to commit a crime in a digital environment without leaving behind any traces at all. Secondly, there was no agreement on the question whether digital evidence may lead to unfair court trials. Some participants reported that the defense lawyers may find it difficult to question the validity of digital evidence, when the prosecutor is not willing to hand over the raw data..

All in all, due to its interdisciplinary composition, the discussion group provided an environment for a fruitful exchange of technical as well as legal aspects.

4.2 Cross-border cloud investigations

York Yannikos

License  Creative Commons BY 3.0 Unported license
© York Yannikos

The topic of this discussion was identifying problems and potential solutions regarding digital forensic investigations in cloud environments. Since the technology used in cloud computing poses new questions for experts in digital forensics and law, several aspects were discussed from a legal and a technical point of view.

The following topical question fields were discussed:

1. What are the currently used methods to acquire digital evidence stored in the cloud? What are the legal requirements?
2. Is it possible to locate evidence data stored in the cloud in a reliable way? What could be done if evidence data could not be located (i.e., the country/countries are unknown where the servers are located on which the data is stored)?

3. What methods/technology could be used in the future to allow a lawful interception of cloud environments?
4. Is it legal to access potential evidence data stored in the cloud through the data owner's computer, e.g., during a house search?

The following sections summarize the results taken from the discussion.

Acquisition of digital evidence in the cloud

By now, cloud data is typically acquired by first localizing where the data is stored in order to go there and seize the corresponding storage hardware or clone any relevant virtual machines. After that, a forensic analysis of the seized storage/VMs is performed. Obviously, without a successful localization of the data and the corresponding hardware the data has to be acquired in a different way.

Currently, existing laws typically regulate access to physical assets or devices where assets are stored. No laws exist which deal with cloud data acquisition in forensic investigations; the consensus of the discussion group was however that such laws are urgently needed.

Localization of digital evidence in the cloud

Currently no standardized techniques/approaches are available for forensic investigators which would allow a reliable data localization in the cloud. Cloud storage technology makes it not only difficult to identify and locate the hardware where specific data is stored, but even the countries where the hardware is located could be hard to tell.

One approach to access cloud data that could not be localized is by using credentials provided by the cloud service provider (CSP). However, this requires the CSP to cooperate and becomes a difficult and/or very time-consuming task e.g., if the CSP is based in another country with different legislation.

Another approach is wiretapping the communication between user and CSP. However, this could require data decryption or reverse engineering of the used protocol which drastically increases the difficulty to access the data.

Future evidence acquisition in the cloud

Theoretically, specific APIs allowing a lawful interception of cloud data could be implemented by the CSPs, but in practice such APIs currently do not exist. However, it is very likely that some kind of forensic API will be provided by the CSPs in the future, but corresponding legal regulations have yet to be defined.

When forensic APIs are available at some point in the future, it is very important to implement controls which strictly regulate and document any use of the APIs in order to prevent abuse and ensure that they are used only for lawful interception.

Legal issues of evidence acquisition in the cloud

Two different opinions were stated within the discussion group:

From the legal point of view, there is no specific law regulating data access in cloud environments. Therefore, existing laws which currently regulate access to assets residing in specific countries have to be applied. As a consequence, access to data in cloud environments is legally permitted only if the location of the data is known and if a legal agreement which permits access to the data exists with the country where the data is localized. This holds also for data which could be possible evidence.

However, from a technical point of view, accessing data stored in the cloud by using the data owner’s computer should be allowed, e.g. when the data owner is currently logged into his cloud storage account. Technically, it makes not much sense to restrict access to data stored in the cloud just because the storage location of the cloud data is not known, since the data itself is typically accessed/processed at the data owner’s computer, much like the data stored on the local hard drive.

4.3 Forensics vs. Privacy

Felix Freiling (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license

© Felix Freiling

Joint work of Gerrit Hornung, Nicolas von zur Mühlen, Dennis Heinson, Jan Ellermann, Hannes Federrath, Andreas Dewald

The discussion group quickly realized that the proper title of this group should have been “Forensics and Data Protection”. This is more general and points to many critical aspects of digital investigations of which the community is largely unaware. This is slowly changing, as indicated by Jan Ellerman, who stated that at Europol, forensic workplaces are filmed and the filmed material is later reviewed to check for proper evidence handling.

Data protection concerns both the collection of evidence and the use of evidence in court. The main legal restriction, at least in Germany, is the *protection of the core area* (“Kernbereichsschutz”). It is constitutionally required to protect the core area of privacy in all procedures. For example, in so-called online searches (collection of evidence through a trojan on the computer of a suspect) a data protection officer, a lawyer and one other person must check and possibly delete data which is considered to belong to the core area.

In digital forensics, the principle of data avoidance is important not only because of the huge amount of data collected by law enforcement. But data avoidance is in conflict with data retention. It was discussed, whether this is part of forensic computing? Another issue of data protection is the sharing of data across borders, e.g. with states that are assumed not to respect human rights in a way that we would expect.

The group did not finish to discuss all relevant problems. More specific questions were to be discussed if group continues (which it unfortunately did not). As a bottom line, all participants agreed that it is important to raise data protection awareness in digital investigations.

4.4 Offensive Countermeasures

Thomas Schreck (SIEMENS CERT München, DE), Michael Gruhn (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license

© Thomas Schreck, Michael Gruhn

Joint work of Felix Freiling, Thomas Schreck, Radim Polcak, Vaclav Stupka, Michael Gruhn, Tatiana Tropina

The discussion was based on five scenarios:

1. Hacking into systems to identify the attacker

Attackers are using several techniques to hide their identification. One way is to hide themselves behind proxy chains provided by different hosting providers or compromised computer systems and located in different countries. In order to identify the attacker, an investigator must follow the chain of proxies back. One way is to break into each system until the attacker's system is identified. There the analyst is able to collect information about the attack and person. A different, often unpractical way, for an investigator is to subpoena her way through the proxy chain. However, because of different jurisdictions this is tedious, sometimes even impossible, and most of the time takes too long to catch an attacker red handed or even at all.

2. Stealing back data an attacker gathered, e.g. via a trojan

Criminals are using so-called dropzone systems to collect stolen information, such as user credentials, online banking credentials, and documents. These dropzones can be readily identified by analyzing the malicious software. However in order to "get the data back", i.e., determine what data has been compromised and act accordingly, it is often necessary to exploit vulnerabilities within the dropzone software to get access to the system. However, again the legal basis for this is unclear, because especially private investigators would be using unauthorized access in violation of some law. Further again the problem of jurisdiction makes this approach difficult to judge legally.

3. Sinkholing malicious systems

IT security researchers are using a technique called "sinkholing" to redirect malicious traffic originally sent to a so-called command and control (C&C) server, to a sinkhole, i.e. a system that analyzes and rejects bad traffic. However, legally this could, in some jurisdictions, be violating telecommunication laws, because the original traffic is diverted, i.e., intercepted.

4. DoS against attacker's controlled systems

The most common attack type on the Internet are denial of service (DoS) attacks. In a DoS attack a malicious entity overloads the service provider with bogus request so legitimate users are denied access to the service. A very simple idea to interrupt the operations of attackers is to use a DoS attack against them. However, there is no explicit legal basis for self-defense on the Internet, hence, such actions, especially when interrupting the service of infrastructure not belonging to the attacker, e.g., intermediate routing networks between the attacker and the investigator, can make these actions just as illegal as the operations of the attacker.

5. Blacklisting and blocking of malicious systems

Another simple way to stop malicious operations is to blacklist and block the systems used to facilitate them. An example for this are the various blacklists for web servers sending spam emails. However, sometimes spammers use legitimate mail servers or networks of hosting providers for their activities. It thus often happens that the mail or hosting providers IP range is blacklisted, even though the mail or hosting provider has already removed the malicious user from their service. This can lead to DoS against the mail or hosting provider. Legally there are no clear guidelines to whether or not a service provider, here the mail providers receiving mail from a blacklisted system, has the right to freely choose whom he provides service to or not. However, this clearly violates net neutrality.

The conclusion of the discussion was that many of the new and often offensive methods that investigators can deploy are not very well covered by the current laws, and depending on how the laws are interpreted and what jurisdictions are competent some of these actions may or may not be legal. However, clear consensus could not be reached. Practical forensic investigators voiced their pledge for clearer laws on the matter.

4.5 Internal investigations

Christian Hawellek (Leibniz Universität Hannover, DE), Michael Gruhn (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Christian Hawellek, Michael Gruhn

Joint work of Gerrit Horning, Christian Hawellek, Dennis Heinson, Thomas Schreck, Radim Polcak, Andreas Dewald, Sven Kälber, Michael Gruhn

Main reference Article 82 of the proposed EU General Data Protection Regulation

In this discussion group Article 82 of the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) was discussed.

First the article was read by everyone and immediately the paragraph 1 c) was found to be of particular importance, as it proposes that “processing of accumulated traffic data shall be permitted in particular to ensure data security”, however, the term “data security” is not defined in the terms of Article 82. Instead Article 30 regulates data security, its scope and protection of personal data. However, by going through these definitions the definition of profiling was confusing and found to be impractical, because searching log files for heuristics, such as attack signatures, needs to be done without existing suspicion, but this would be inadmissible profiling according to the proposed regulations.

It was discussed whether this searching for heuristics is really profiling. The proposal defines profiling in Art. 4 Par. 3 a). It states that “‘profiling’ means any form of automated processing of personal data intended to evaluate certain personal aspects relating to a natural person or to analyse or predict in particular that natural person’s performance at work, economic situation, location, health, personal preferences, reliability or behaviour”. By this definition technical profiling even without personal profiling is hardly possible, as is any form of data aggregation.

Thus it was discussed whether Art. 82 Par. 1 c) is *lex specialis* against Par. 1 a) and whether technical detection was sufficiently considered in the proposal. One possible argument that makes profiling legal under Art. 4 is when the intentional element is missing, i.e., the data is not intentionally aggregated to gain personal information, but to investigate attacks. For example, an investigator collects all emails, data on a system and browser history, then performs a key word search on this data. In this circumstance it would be a compliance investigation search and Par. 1 c) would apply.

Discussing the impact of these definitions it was agreed that the definition of profiling is too broad. Depending on how the definition is interpreted it may apply to all investigations of IT security teams, as they are always automatically process data involving certain personal aspects. However, reading the definition thus that a personal profile needs to be the result of profiling may exclude those cases in which only a technical analysis takes place for IT security reasons.

Because of time constraints further definitions, e.g., for “freely given consent” were not discussed.

However, the question whether profiling actually exists in companies was briefly discussed. One such example is Google. Some years ago they tried to predict how likely an employee would be to quit from their web browsing habits to be able to react on this proactively. Pseudonymization might be a way to avoid the risk of such profiling being applied, however, pseudonymization is not possible with unstructured data.

The next issue that was discussed was the term “appropriate” in Par. 1 c). Par. 1 c) deals with deadlines for deletion of gathered data. There used to be an allowance for judgment depending on the circumstances of the case. However, in the proposal it is unclear whether deadlines for deletion refer to the individual investigations or not. This is problematic because the provision refers to “processing” and to “investigating” both the same. It is also unclear whether or not a general information such as “your email traffic will be under surveillance” is sufficient for compliance, because this would create some sort of general knowledge.

The proposal further does not cover routine checks, e.g. checking employees computers for malware such as trojans. A particular suspicion is always required if personal data is concerned. A highly problematic scenario is that general surveillance may exist in a company, from which suspicion is raised about a particular incidence, on the basis of which an investigation is initiated. However, it could also be that such information is used to retroactively justify the investigation by not revealing that only due to the surveillance measure the suspicion has been created in the first place. Another concern is proportionality as it applies to both the process of collection as such and the nature and extent of the data collection it remains unclear whether “data collection” in the second half of the Par. 1 c) refers to the process or the outcome. It is also unclear what “nature of data collection” means. Is the purpose the nature or is the way in which the data is collected the nature?

Par. 1 c) (a) further states “the investigation shall be carried out by the competent authority” but it is undefined whether “investigation” and “authority” refer to public authorities’ investigation or private investigations. Whereas the terms very much indicate so, it would be difficult to put this in context with the alternative of “serious dereliction of duty in the employment” context.

Par. 1 d) again raised terminology discussions between forensic practitioners and the legal professionals. The questions discussed were: What does data security mean? What is accumulated traffic data? Does it mean anonymous? Why are “Internet” and “email” mentioned separately? To include email in private networks? Is traffic data the same as content data? Technically the paragraph states that only meta data shall be analyzed. But this is not effective for data security. Cannot not be used for preventive measures (Lacuna).

Besides the main topic also the wording of laws in general was discussed spearheaded by the question why is the word “must” sometimes changed into “shall”? This could have historic reasons and possibly to emphasize the normative element. But does not change the meaning because “shall” can be replaced with “must” without changing the meaning of the law text.

To summarize, the following problems have been carved out during the discussions that are not very well covered by the proposed regulation:

- content data (what is it and is it personal or not?)
- preventive measures (monitoring for threat detection)
- wording (internet, data security, profiling)
- external attackers (lacuna)
- server forensics (exemption missing for fragments of personal data).

Participants

- Andreas Dewald
Univ. Erlangen-Nürnberg, DE
- Jan Ellermann
Europol, NL
- Hannes Federrath
Universität Hamburg, DE
- Felix C. Freiling
Univ. Erlangen-Nürnberg, DE
- Michael Gruhn
Univ. Erlangen-Nürnberg, DE
- Christian Hawellek
Leibniz Univ. Hannover, DE
- Dennis Heinson
Hamburg, DE
- Dominik Herrmann
Universität Hamburg, DE
- Gerrit Hornung
Universität Passau, DE
- Sven Kälber
Univ. Erlangen-Nürnberg, DE
- Stefan Kiltz
Universität Magdeburg, DE
- Volker Krummel
Wincor-Nixdorf International
GmbH – Paderborn, DE
- Radim Polcák
Masaryk University, CZ
- Thomas Schreck
Siemens – München, DE
- Erich Schweighofer
Universität Wien, AT
- Tobias Singelstein
Freie Universität Berlin, DE
- Vaclav Stupka
Masaryk University – Brno, CZ
- Tatiana Tropina
MPI für Strafrecht –
Freiburg, DE
- Nicolas von zur Mühlen
MPI für Strafrecht –
Freiburg, DE
- York Yannikos
Fraunhofer SIT – Darmstadt, DE
- Riha Zdenek
Masaryk University – Brno, CZ

