



Volume 9, Issue 1, January 2019

| | |
|--|-----|
| Joint Processing of Language and Visual Data for Better Automated Understanding (Dagstuhl Seminar 19021) <i>Marie-Francine Moens, Lucia Specia, and Tinne Tuytelaars</i> | 1 |
| ALogics for Dependence and Independence (Dagstuhl Seminar 19031) <i>Erich Grädel, Phokion G. Kolaitis, Juha Kontinen, and Heribert Vollmer</i> | 28 |
| Conditional Logics and Conditional Reasoning: New Joint Perspectives (Dagstuhl Seminar 19032) <i>Guillaume Aucher, Paul Egré, Gabriele Kern-Isberner, and Francesca Poggiolesi</i> | 47 |
| New Horizons in Parameterized Complexity (Dagstuhl Seminar 19041) <i>Fedor V. Fomin, Dániel Marx, Saket Saurabh, and Meirav Zehavi</i> | 67 |
| Practical Yet Composably Secure Cryptographic Protocols (Dagstuhl Seminar 19042) <i>Jan Camenisch, Ralf Küsters, Anna Lysyanskaya, and Alessandra Scafuro</i> | 88 |
| Data Structures for the Cloud and External Memory Data (Dagstuhl Seminar 19051) <i>Gerth Stølting Brodal, Ulrich Carsten Meyer, Markus E. Nebel, and Robert Sedgewick</i> | 104 |
| Computational Methods for Melody and Voice Processing in Music Recordings (Dagstuhl Seminar 19052) <i>Meinard Müller, Emilia Gómez, and Yi-Hsuan Yang</i> | 125 |

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/dagpub/2192-5283>

Publication date

August, 2019

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 DE license (CC BY 3.0 DE).



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

- Gilles Barthe
- Bernd Becker
- Daniel Cremers
- Stephan Diehl
- Reiner Hähnle
- Lynda Hardman
- Oliver Kohlbacher
- Bernhard Mitschang
- Bernhard Nebel
- Albrecht Schmidt
- Wolfgang Schröder-Preikschat
- Raimund Seidel (*Editor-in-Chief*)
- Emanuel Thomé
- Heike Wehrheim
- Verena Wolf
- Martina Zitterbart

Editorial Office

Michael Wagner (*Managing Editor*)
Jutka Gasiorowski (*Editorial Assistance*)
Dagmar Glaser (*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
<http://www.dagstuhl.de/dagrep>

Digital Object Identifier: 10.4230/DagRep.9.1.i

Report from Dagstuhl Seminar 19021

Joint Processing of Language and Visual Data for Better Automated Understanding

Edited by

Marie-Francine Moens¹, Lucia Specia², and Tinne Tuytelaars³

1 KU Leuven, BE, sien.moens@cs.kuleuven.be

2 Imperial College London, GB, lspecia@gmail.com

3 KU Leuven, BE, tinne.tuytelaars@esat.kuleuven.be

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19021 “Joint Processing of Language and Visual Data for Better Automated Understanding”. It includes a discussion of the motivation and overall organization, the abstracts of the talks, and a report of each working group.

Seminar January 6–11, 2019 – <http://www.dagstuhl.de/19021>

2012 ACM Subject Classification Computing methodologies → Natural language processing, Computing methodologies → Natural language generation, Computing methodologies → Computer vision tasks, Computing methodologies → Scene understanding, Computing methodologies → Image representations

Keywords and phrases multimodal learning, representation learning, computer vision, natural language processing, machine learning

Digital Object Identifier 10.4230/DagRep.9.1.1

Edited in cooperation with Guillem Collell

1 Executive Summary

Marie-Francine Moens

Lucia Specia

Tinne Tuytelaars

License  Creative Commons BY 3.0 Unported license
© Marie-Francine Moens, Lucia Specia, and Tinne Tuytelaars

The joint processing of language and visual data has recently received a lot of attention. This emerging research field is stimulated by the active development of deep learning algorithms. For instance, deep neural networks (DNNs) offer numerous opportunities to learn mappings between the visual and language media and to learn multimodal representations of content. Furthermore, deep learning recently has become a standard approach for automated image and video captioning and for visual question answering, the former referring to the automated description of images or video with descriptions in natural language sentences, the latter to the automated formulation of an answer in natural language to a question in natural language about an image.

Apart from aiding image understanding and the indexing and search of image and video data through the natural language descriptions, the field of jointly processing language and visual data builds algorithms for grounded language processing where the meaning of



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

Joint Processing of Language and Visual Data for Better Automated Understanding,

Dagstuhl Reports, Vol. 9, Issue 1, pp. 1–27

Editors: Marie-Francine Moens, Lucia Specia, and Tinne Tuytelaars



DAGSTUHL
REPORTS Dagstuhl Reports

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

natural language is based on perception and/or actions in the world. Grounded language processing contributes to automated language understanding and machine translation of language. Recently, it has been shown that visual data provide world and common-sense knowledge that is needed in automated language understanding.

Joint processing of language and visual data is also interesting from a theoretical point of view for developing theories on the complementarity of such data in human(-machine) communication, for developing suitable algorithms for learning statistical knowledge representations informed by visual and language data, and for inferencing with these representations.

Given the current trend and results of multimodal (language and vision) research, it can be safely assumed that the joint processing of language and visual data will only gain in importance in the future. During the seminar we have discussed theories, methodologies and real-world technologies for joint processing of language and vision, particularly in the following research areas:

- Theories of integrated modelling and representation learning of language and vision for computer vision and natural language processing tasks;
- Explainability and interpretability of the learned representations;
- Fusion and inference based on visual, language and multimodal representations;
- Understanding human language and visual content;
- Generation of language and visual content;
- Relation to human learning;
- Datasets and tasks.

The discussions have attempted to give an answer to the following research questions (a non-exhaustive list):

- Which machine learning architectures will be best suited for the above tasks?
- How to learn multimodal representations that are relational and structured in nature to allow a structured understanding?
- How to generalize to allow recognitions that have few or zero examples in training?
- How to learn from limited paired data but exploiting monomodal models trained on visual or language data?
- How to explain the neural networks when they are trained for image or language understanding?
- How to disentangle the representations: factorization to separate the different factors of variation and discovering of their meaning?
- How to learn continuous representations that describe semantics and that integrate world and common-sense knowledge?
- How to reason with the continuous representations?
- How to translate to another modality?
- What would be effective novel evaluation metrics?

This Dagstuhl Seminar has brought together an interdisciplinary group of researchers from computer vision, natural language processing, machine learning and artificial intelligence to discuss the latest scientific realizations and to develop a roadmap and research agenda.

2 Table of Contents

Executive Summary

| | |
|---|---|
| <i>Marie-Francine Moens, Lucia Specia, and Tinne Tuytelaars</i> | 1 |
|---|---|

Overview Talks

| | |
|---|----|
| Science in Computer Science | |
| <i>Andrei Barbu</i> | 5 |
| Grounded Language Learning in Virtual Environments | |
| <i>Stephen Clark</i> | 5 |
| RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes | |
| <i>Aykut Erdem</i> | 6 |
| Language Based Image Manipulation | |
| <i>Erkut Erdem</i> | 6 |
| Learning from Multilingual Multimodal Data | |
| <i>Desmond Elliott</i> | 7 |
| Women also Snowboard: Overcoming Bias in Image Captioning | |
| <i>Lisa Anne Hendricks</i> | 7 |
| Visual Context for Verb Sense Disambiguation | |
| <i>Frank Keller</i> | 8 |
| Countering Language Drift through Grounding / Reinterpreting Wittgenstein | |
| <i>Douwe Kiela</i> | 8 |
| Mining for Meaning: From Vision to Language through Multiple Networks Consensus | |
| <i>Marius Leordeanu</i> | 9 |
| Describing Similarities and Differences in Related Videos | |
| <i>Florian Metze</i> | 9 |
| Visual Dialogue without Vision or Dialogue | |
| <i>Siddharth Narayanaswamy</i> | 10 |
| Conversational Mobile Robots: Integrating Vision, Language, and Planning | |
| <i>Jean Oh</i> | 10 |
| Grounding Semantic Roles in Images | |
| <i>Carina Silberer</i> | 11 |
| Pervasive Attention: 2D CNNs for Sequence-to-Sequence Prediction | |
| <i>Jakob Verbeek</i> | 11 |

Working Groups

| | |
|--|----|
| Representations | |
| <i>Tinne Tuytelaars, Vera Demberg, David Hogg, Dietrich Klakow, Chiraag Lala, Jindrich Libovický, Pranava Madhyastha, Siddharth Narayanaswamy, Bernt Schiele</i> | 12 |
| Visual and Language Understanding | |
| <i>Marie-Francine Moens, Luisa Coheur, Stephen Clark, Erkut Erdem, Anette Frank, Jean Oh</i> | 14 |

| | |
|---|----|
| Challenges in Language Generation Applications <i>Lucia Specia, Loïc Barrault, Thales Bertaglia, Erkut Erdem, Lisa Anne Hendricks, Pecina Pavel, Florian Metze, Jean Oh</i> | 17 |
| Modeling Human Learning <i>Raffaella Bernardi, Zeynep Akata, Andrei Barbu, Ozan Caglayan, Stephen Clark, Guillem Collell, Desmond Elliott, Raquel Fernandez, Orhan Firat, Stella Frank, Frank Keller, Douwe Kiela, Pecina Pavel, David Vernon, Josiah Wang</i> | 21 |
| Explainability <i>Tinne Tuytelaars, Luisa Coheur, Vera Demberg, Lisa Anne Hendricks, Dietrich Klakow, Jindrich Libovický, Pranava Madhyastha, Marie-Francine Moens, Siddharth Narayanaswamy, Bernt Schiele</i> | 23 |
| Tasks: Creating Simulated Worlds from Existing Media <i>David Hogg, Raffaella Bernardi, Desmond Elliott, Raquel Fernandez, Stella Frank, Marius Leordeanu, Jean Oh, Pavel Pecina, Lucia Specia, Jakob Verbeek, David Vernon</i> | 24 |
| Tasks and Datasets for Vision and Language <i>Stephen Clark, Zeynep Akata, Andrei Barbu, Loïc Barrault, Raffaella Bernardi, Ozan Caglayan, Aykut Erdem, Erkut Erdem, Orhan Firat, Anette Frank, Stella Frank, David Hogg, Frank Keller, Douwe Kiela, Chiraag Lala, Marius Leordeanu, Florian Metze, Lucia Specia</i> | 25 |
| Participants | 27 |

3 Overview Talks

3.1 Science in Computer Science

Andrei Barbu (MIT – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Andrei Barbu

Joint work of Andrei Barbu, Yevgeni Berzak, David Mayo, Boris Katz

The methodology used to run experiments in computer science, for determining the performance of systems and humans, has been the same for several decades despite the huge advances in machine learning techniques. In particular, the ability of deep networks to effectively exploit unforeseen correlations in datasets means that many datasets are no longer good predictors of real-world performance. We demonstrate this with a new methodology to collect large-scale image datasets on Mechanical Turk while controlling for biases. De-biasing data is standard in other sciences and we believe computer science should follow: put simply, collecting images at random from some source does not guarantee those images are not highly biased. In another domain, we demonstrate how human performance has been significantly overstated for syntactic parsing problems and demonstrate that existing systems are overfitting to the particular biased methodology used to annotate parsing datasets. As datasets grow in importance in machine learning, and computer science in general, it is time that we adopt the lessons from other sciences: stringent controls, independently collecting test and training sets, and characterizing human performance at scale to create baselines for machines and to discover new biases.

3.2 Grounded Language Learning in Virtual Environments

Stephen Clark (Google DeepMind – London, GB)

License © Creative Commons BY 3.0 Unported license
© Stephen Clark

Joint work of Stephen Clark, Felix Hill, Karl Moritz Hermann, Phil Blunsom

Main reference Felix Hill, Karl Moritz Hermann, Phil Blunsom, Stephen Clark: “Understanding Grounded Language Learning Agents”, CoRR, Vol. abs/1710.09867, 2017.

URL <https://arxiv.org/abs/1710.09867>

Images provide a means of grounding natural language expressions in another modality. However, there are limits to what images can provide in this regard, including a lack of state change and a lack of agent interaction. In this talk we describe work in grounding language in the actions of an embodied agent, where the environment of the agent is a simulated virtual world. We focus on some limitations of the current work, and discuss how far such an approach could take us in the goal of developing intelligent agents in both a virtual and the real world.

3.3 RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes

Aykut Erdem (Hacettepe University – Ankara, TR)

License © Creative Commons BY 3.0 Unported license
© Aykut Erdem

Joint work of Semih Yagcioglu, Aykut Erdem, Ekut Erdem, Nazli Ikizler-Cinbis
Main reference Semih Yagcioglu, Aykut Erdem, Ekut Erdem, Nazli Ikizler-Cinbis: “RecipeQA: A Challenge Dataset for Multimodal Comprehension of Cooking Recipes”, In Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, Brussels, Belgium, Oct 31-Nov 4, 2018, pp. 1358–1368, ACL, 2018.
URL <https://aclweb.org/anthology/D18-1166>

Understanding and reasoning about cooking recipes is a fruitful research direction towards enabling machines to interpret procedural text. In this talk, we present RecipeQA, a dataset for multimodal comprehension of cooking recipes. It comprises of approximately 20K instructional recipes with multiple modalities such as titles, descriptions and aligned set of images. With over 36K automatically generated question-answer pairs, we design a set of comprehension and reasoning tasks that require joint understanding of images and text, capturing the temporal flow of events and making sense of procedural knowledge. Our preliminary results indicate that RecipeQA serves as a challenging test bed and an ideal benchmark for evaluating machine comprehension systems. The data and leaderboard are available at <https://hucvl.github.io/recipeqa/>

3.4 Language Based Image Manipulation

Erkut Erdem (Hacettepe University – Ankara, TR)

License © Creative Commons BY 3.0 Unported license
© Erkut Erdem

Joint work of Levent Karacan, Zeynep Akata, Aykut Erdem, Erkut Erdem
Main reference Levent Karacan, Zeynep Akata, Aykut Erdem, Erkut Erdem: “Manipulating Attributes of Natural Scenes via Hallucination”, CoRR, Vol. abs/1808.07413, 2018.
URL <http://arxiv.org/abs/1808.07413>

Recently, much progress has been made towards realistic image synthesis. In particular, different flavors and improved versions of Generative Adversarial Networks (GANs) have achieved impressive results along this direction. Using GANs as a backbone, we present our efforts on language based image manipulation. In our first study, we explore building a two-stage framework for enabling users to directly manipulate high-level attributes of a natural scene. The key to our approach is a deep generative network which can hallucinate images of a scene as if they were taken at a different season (e.g., during winter), weather condition (e.g., in a cloudy day) or time of the day (e.g., at sunset). In our second work, we present a novel approach for language conditioned editing of fashion images. Our approach employs a GAN-based architecture which allows the users to edit an outfit image by feeding in different descriptions to generate new outfits.

3.5 Learning from Multilingual Multimodal Data

Desmond Elliott (University of Copenhagen – DK)

License © Creative Commons BY 3.0 Unported license
© Desmond Elliott

Joint work of Elliott Desmond, Ákos Kádár

Main reference Desmond Elliott, Ákos Kádár: “Imagination Improves Multimodal Translation.” In Proc. of the Eighth International Joint Conference on Natural Language Processing, EMNLP 2017, Taipei, Taiwan, November 27-December 1, 2017, (Volume 1: Long Papers) Vol. 1, pp. 130–141, ACL 2017.

URL <https://aclweb.org/anthology/I17-1014>

We speak about two perspectives on learning from multilingual and multimodal data. In the language generation setting of multimodal machine translation, we discuss whether we should use visual representations as an input variable, or as a variable that the model learns to predict. In the image–sentence retrieval setting, we discuss experiments on when it is useful to train with multilingual annotations, as opposed to monolingual annotations.

3.6 Women also Snowboard: Overcoming Bias in Image Captioning

Lisa Anne Hendricks (University of California – Berkeley, US)

License © Creative Commons BY 3.0 Unported license
© Lisa Anne Hendricks

Joint work of Lisa Anne Hendricks, Kaylee Burns, Kate Saenko, Trevor Darrell, Anna Rohrbach

Main reference Lisa Anne Hendricks, Kaylee Burns, Kate Saenko, Trevor Darrell, Anna Rohrbach: “Women Also Snowboard: Overcoming Bias in Captioning Models”, in Proc. of the Computer Vision – ECCV 2018 – 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part III, Lecture Notes in Computer Science, Vol. 11207, pp. 793–811, Springer, 2018.

URL https://doi.org/10.1007/978-3-030-01219-9_47

Most machine learning methods are known to capture and exploit biases of the training data. While some biases are beneficial for learning, others are harmful. Specifically, image captioning models tend to exaggerate biases present in training data (e.g., if a word is present in 60% of training sentences, it might be predicted in 70% of sentences at test time). This can lead to incorrect captions in domains where unbiased captions are desired, or required, due to over-reliance on the learned prior and image context. In this work we investigate generation of gender-specific caption words (e.g., man, woman) based on the person’s appearance or the image context. We introduce a new Equalizer model that ensures equal gender probability when gender evidence is occluded in a scene and confident predictions when gender evidence is present. The resulting model is forced to look at a person rather than use contextual cues to make a gender-specific predictions. The losses that comprise our model, the Appearance Confusion Loss and the Confident Loss, are general, and can be added to any description model in order to mitigate impacts of unwanted bias in a description dataset. Our proposed model has lower error than prior work when describing images with people and mentioning their gender and more closely matches the ground truth ratio of sentences including women to sentences including men. We also show that unlike other approaches, our model is indeed more often looking at people when predicting their gender.

3.7 Visual Context for Verb Sense Disambiguation

Frank Keller (University of Edinburgh, GB)

License  Creative Commons BY 3.0 Unported license
© Frank Keller

Joint work of Frank Keller, Gella Spandana, Elliott Desmond

Main reference Spandana Gella, Frank Keller, Mirella Lapata: “Disambiguating Visual Verbs”, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 41(2), pp. 311–322, 2019.

URL <https://doi.org/10.1109/TPAMI.2017.2786699>

Every day millions of images are uploaded on the web. To process images at such a large scale it is important to build automatic image understanding systems. A step towards understanding the content of an image is to be able to recognize actions (or verbs) depicted in the image. This type of image understanding can then be integrated with natural language processing to build systems that interact with humans for tasks such as image retrieval. In this talk, we present models for integrating visual and textual contexts for: (i) Verb Classification: automatically identifying verbs that denote actions depicted in images; (ii) Visual Sense Disambiguation: fine-grained analysis of how visual context can help disambiguate different meanings of verbs; (iii) Multilingual Sense Disambiguation: using visual sense disambiguation across languages to benefit tasks such as machine translation.

3.8 Countering Language Drift through Grounding / Reinterpreting Wittgenstein

Douwe Kiela (Facebook – New York, US)

License  Creative Commons BY 3.0 Unported license
© Douwe Kiela

The distributional hypothesis (Harris, 1952; Firth, 1957) has come to be one of the corner stones of modern natural language processing (NLP). It is the foundation upon which word embeddings are built, for instance. Oftentimes, the distribution hypothesis is seen as an incarnation of Wittgenstein’s famous “meaning is use” paradigm. In this talk, we argue that that conception of Wittgenstein as “the godfather of the distributional hypothesis” is misguided, and that Wittgenstein has rather different lessons to teach us.

The distributional hypothesis has two big issues. First, it forces us to define symbols only in terms of other symbols, which leads to the well-studied grounding problem. Second, and this is less well-studied, it presumes a passive observational stance towards language, where we just observe “the company that words keep”. We argue that Wittgenstein’s actual theory of meaning and language does not suffer from either of these issues, and that we should take this account of language more seriously. The resultant interpretation of Wittgenstein leads to a new research program for true natural language understanding, centering around active language usage in “multi-agent grounded language games”. We finish with some quick examples of research we have done at FAIR that goes in that direction.

3.9 Mining for Meaning: From Vision to Language through Multiple Networks Consensus

Marius Leordeanu (University Politehnica of Bucharest, RO)

License © Creative Commons BY 3.0 Unported license
© Marius Leordeanu

Joint work of Leordeanu, Marius; Iulia, Duta; Andrei Liviu, Nicolicioiu, Simion-Vlad, Bogolin

Main reference Iulia Duta, Andrei Liviu Nicolicioiu, Simion-Vlad Bogolin, Marius Leordeanu: “Mining for meaning: from vision to language through multiple networks consensus”, In Proc. of the 29th British Machine Vision Conference, BMVC 2018, Newcastle, GB, Sep 3-6, 2018.

URL <http://bmvc2018.org/contents/papers/1031.pdf>

Describing visual data into natural language is a very challenging task, at the intersection of computer vision, natural language processing and machine learning. Language goes well beyond the description of physical objects and their interactions and can convey the same abstract idea in many ways. It is both about content at the highest semantic level as well as about fluent form. Here we propose an approach to describe videos in natural language by reaching a consensus among multiple encoder-decoder networks. Finding such a consensual linguistic description, which shares common properties with a larger group, has a better chance to convey the correct meaning. We propose and train several network architectures and use different types of image, audio and video features. Each model produces its own description of the input video and the best one is chosen through an efficient, two-phase consensus process. We demonstrate the strength of our approach by obtaining state of the art results on the challenging MSR-VTT dataset.

3.10 Describing Similarities and Differences in Related Videos

Florian Metze (Carnegie Mellon University – Pittsburgh, US)

License © Creative Commons BY 3.0 Unported license
© Florian Metze

Joint work of Jindrich Libovický, Shruti Palaskar, Spandana Gella, Florian Metze

Main reference Jindrich Libovický, Shruti Palaskar, Spandana Gella, Florian Metze: “Multimodal abstractive summarization of opendomain videos”, In NeurIPS Workshop on Visually Grounded Interaction and Language (ViGIL), 32nd Conference on Neural Information Processing Systems, NIPS 2018, Montréal, Canada. Dec 3-8, NIPS, 2018.

URL <https://nips2018vigil.github.io/static/papers/accepted/8.pdf>

Multimodal and abstractive summarization of open-domain videos requires summarizing the contents of an entire video in a few short sentences, while fusing information from multiple modalities, in our case video and audio (or text). Different from traditional news summarization, the goal is less to “compress” text information only, but to provide a fluent textual summary of information that has been collected and fused from different source modalities. In this talk, we introduce the task of abstractive summarization for open-domain videos, we show how a sequence-to-sequence model with hierarchical attention can integrate information from different modalities into a coherent output, and present pilot experiments on the How2 corpus of instructional videos. We also present a new evaluation metric for this task called Content F1 that measures semantic adequacy rather than fluency of the summaries, which is covered by ROUGE and BLEU like metrics.

3.11 Visual Dialogue without Vision or Dialogue

Siddharth Narayanaswamy (University of Oxford, GB)

License © Creative Commons BY 3.0 Unported license
© Siddharth Narayanaswamy

Joint work of Siddharth Narayanaswamy, Daniela Massiceti, Puneet Dokania, Philip Torr

Main reference Daniela Massiceti, Puneet K. Dokania, N. Siddharth, Philip H. S. Torr: “Visual Dialogue without Vision or Dialogue”, CoRR, Vol. abs/1812.06417, 2018.

URL <https://arxiv.org/abs/1812.06417>

We characterize some of the quirks and shortcomings in the exploration of Visual Dialogue (VD) – a sequential question-answering task where the questions and corresponding answers are related through given visual stimuli. Using an embarrassingly simple method based on Canonical Correlation Analysis (CCA) on the standard dataset gets near state-of-the-art performance for some standard metric. In contrast to current complex and over-parametrized architectures that are both compute and time intensive, this method ignores the visual stimuli, ignores the sequencing of dialogue, does not need gradients, uses off-the-shelf feature extractors, has at least an order of magnitude fewer parameters, and learns in practically no time. These results are indicative of issues in current approaches to Visual Dialogue relating particularly to implicit dataset biases, under-constrained task objectives, and over-constrained evaluation metrics.

3.12 Conversational Mobile Robots: Integrating Vision, Language, and Planning

Jean Oh (Carnegie Mellon University – Pittsburgh, US)

License © Creative Commons BY 3.0 Unported license
© Jean Oh

Joint work of Jean H. Oh, Arne Suppé, Felix Duvallet, Abdeslam Boularias, Luis E. Navarro-Serment, Martial Hebert, Anthony Stentz, Jerry Vinokurov, Oscar J. Romero, Christian Lebiere, Robert Dean

Main reference Jean H. Oh, Arne Suppé, Felix Duvallet, Abdeslam Boularias, Luis E. Navarro-Serment, Martial Hebert, Anthony Stentz, Jerry Vinokurov, Oscar J. Romero, Christian Lebiere, Robert Dean: “Toward Mobile Robots Reasoning Like Humans”, in Proc. of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., pp. 1371–1379, AAAI Press, 2015.

URL <http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9766>

As robots are envisioned to become ubiquitous in our personal and work environments, there have been growing interests in developing intuitive ways for lay users to interact with such autonomous systems, raising several challenging questions. For instance, can we command robots in natural language? Can robots describe what they observe or explain what they have done or plan to do? How can we train our robots to understand rich semantic context, utilizing a vast amount of sensor data that is available in multiple modalities? In this talk, we discuss various hurdles in addressing these challenges in several robotics application domains including social navigation, autonomous driving, disaster response, and military robots. We will also discuss general limitations of datasets and evaluation metrics in interdisciplinary research and propose alternative directions.

3.13 Grounding Semantic Roles in Images

Carina Silberer (UPF – Barcelona, ES)

License © Creative Commons BY 3.0 Unported license
© Carina Silberer

Joint work of Carina Silberer, Manfred Pinkal

Main reference Carina Silberer, Manfred Pinkal: “Grounding Semantic Roles in Images”, in Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 – November 4, 2018, pp. 2616–2626, Association for Computational Linguistics, 2018.

URL <https://aclanthology.info/papers/D18-1282/d18-1282>

Images of everyday scenes can be interpreted and described in many ways, depending on the perceiver and the context in which the image is presented, where the context may be natural language data or a visual sequence. The interpretation of a (visual) scene is related to the determination of who did what to whom, etc. This may require a joint processing or reasoning with possibly multiple (extra-)linguistic information sources (e.g., text, images).

To facilitate the joint processing over multiple sources, it is desirable to induce representations of texts and visual scenes which do encode this kind of information, and in, essentially, a congruent and generic way. In this talk we present our approach towards this goal: We address the task of visual semantic role labeling (vSRL), and learn frame-semantic representations of images. Our model renders candidate participants as image regions of objects, and is trained towards grounding roles in the regions which depict the corresponding participant. We present experimental results which demonstrate that we can train a vSRL model without reliance on prohibitive image-based role annotations, by utilizing noisy data which we extract automatically from image captions using a linguistic SRL system. Furthermore, the frame-semantic visual representations which our model induces yield overall better results on supervised visual verb sense disambiguation compared to previous work.

3.14 Pervasive Attention: 2D CNNs for Sequence-to-Sequence Prediction

Jakob Verbeek (INRIA – Grenoble, FR)

License © Creative Commons BY 3.0 Unported license
© Jakob Verbeek

Joint work of Jakob Verbeek, Maha Elbayad, Laurent Besacier

Main reference Maha Elbayad, Laurent Besacier, Jakob Verbeek: “Pervasive Attention: 2D Convolutional Neural Networks for Sequence-to-Sequence Prediction”, in Proc. of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018, Brussels, Belgium, October 31 – November 1, 2018, pp. 97-107, Association for Computational Linguistics, 2018.

URL <https://aclweb.org/anthology/K18-1010>

Current state-of-the-art machine translation systems are based on encoder-decoder architectures, that first encode the input sequence, and then generate an output sequence based on the input encoding. Both are interfaced with an attention mechanism that recombines a fixed encoding of the source tokens based on the decoder state. We propose an alternative approach which instead relies on a single 2D convolutional neural network across both sequences. Each layer of our network re-codes source tokens on the basis of the output sequence produced so far. Attention-like properties are therefore pervasive throughout the network. Our model yields excellent results, outperforming state-of-the-art encoder-decoder systems, while being conceptually simpler and having fewer parameters.

4 Working Groups

4.1 Representations

Tinne Tuytelaars (KU Leuven, BE), Vera Demberg (Universität des Saarlandes, DE), David Hogg (University of Leeds, GB), Dietrich Klakow (Universität des Saarlandes, DE), Chiraag Lala (University of Sheffield, GB), Jindrich Libovický (Charles University – Prague, CZ), Pranava Madhyastha (Imperial College London – GB), Siddharth Narayanaswamy (University of Oxford, GB), Bernt Schiele (MPI für Informatik – Saarbrücken, DE)

License  Creative Commons BY 3.0 Unported license

© Tinne Tuytelaars, Vera Demberg, David Hogg, Dietrich Klakow, Chiraag Lala, Jindrich Libovický, Pranava Madhyastha, Siddharth Narayanaswamy, Bernt Schiele

As a group, we had a discussion on *representations*. The way data in general, and images, video and text in particular, are represented defines, to a large extent, what information can (easily) be extracted from the data. Representations are also key when it comes to estimating the similarity between two or more items. With the popularity of deep learning methods, many of the representations used today are learned in a data-driven manner, making it harder to assess what they really capture.

We started the discussion with a definition of the concept of representation, as vision and language communities seemed to have somewhat different views on this. A first important aspect is the *representation scheme* or *format* (e.g., do we represent the data as a vector, matrix, tensor, graph, set or something else?). For most of our discussion, we focused on the case of a vectorial representation. Especially for vision people, this seemed a natural choice, while text people use a larger variety of representations. The main reason to opt for other, more complex formats is to make some of the structure within the data more explicit (e.g., the spatial dimensions of an image captured by a matrix or tensor, or dependencies made explicit in a graph representation). Even within a given format, different representations are possible. Ultimately, these are all the result of some transformations applied to the input data, typically with the aim to remove redundancy, remove noise and highlight relevant information.

Representations can be designed (‘hand-crafted’) or learned from data (typically with neural networks). In the latter case, the representation obtained depends on i) the network architecture (e.g., receptive field), ii) the loss used to train the model (e.g., reconstruction of the input data, semantic classification, etc.), and iii) the training data.

We discussed properties of the “ideal representation”. Desired characteristics (sometimes conflicting) of good representations include:

- *Compactness*;
- *Robustness*: A small change in the input data does not have a big impact on the resulting representation;
- *Disentangled*: Different aspects of the data are stored in different subspaces, e.g., illumination vs. geometry vs. material for objects; ethnicity vs. facial expression vs. age for faces; or intent vs. style vs. language for text messages;
- *Explicit*: Easy to interpret by humans, with different elements being ‘name-able’;
- *Transferable*: A good representation generalizes well beyond the initial training conditions;
- *Probabilistic*: A good representation incorporates information about uncertainty.

Additionally, depending on the context, additional characteristics include

- *Static vs. dynamic*: In case of streaming input data, a good representation gradually changes over time;
- *Task-specific vs. universal*: In some cases it can be beneficial to have a representation that's tuned towards one specific task; yet ideally one can imagine the existence of a universal representation, from which task-specific ones can be derived by projecting on some lower-dimensional subspace;
- *Granularity*: Representations can be considered at different levels of granularity, e.g., objects or scenes in images; or words, sentences or stories in text. For images, there's the additional distinction between 2D and 3D representations.

We had a further discussion about *implicit vs. explicit representations*, and concluded that this distinction is related to the difference between model-driven vs. data-driven approaches, and closely linked to explainability. One interpretation is to measure the degree of explicitness of a representation as the amount of work that is needed (either by a human or a machine) to derive knowledge from it. Making a representation more explicit than often implies imposing more constraints based on prior knowledge (a model).

In the context of language research, both discrete and continuous representations are used. Understanding of language goes beyond semantics (e.g., intent) and explicit representations. Structured representations are an example of more explicit representations – yet opinions seemed to be mixed as to whether this is something we should strive for or maybe not really needed. There was agreement though that explicit symbolic units are probably insufficient to capture the richness of natural language.

In the context of vision research, a trend towards more explicit representations can be observed in the sense that several works aim at incorporating domain knowledge such as geometric constraints in the neural network models. There's also old work explicitly designing neural networks with weights derived from physics or geometry. The work on Generative Query Networks, on the other hand, is an example of a powerful implicit representation.

We continued with a discussion on the interaction between visual and textual representations. Both modalities are complementary, and visual understanding cannot be reduced to just mapping images to words. Language, e.g., in the form of image captions or textual descriptions, seems especially useful to provide compositionality in human learning, leading to more abstract interpretations and better generalization. It can also help to focus attention. At the same time, the two communities think about representations quite differently: while the language community is mostly aiming to deal with the ambiguity in language and stresses the fact that a lot of information is implicit, vision researchers aim for a precise, absolute description, and mostly ignore implicit aspects.

Open research questions include the development of hybrid models combining data-driven learning with more explicit representations (especially on the vision side), representations of the 3D world (vision), better cross-modal representations, and unsupervised machine translation using statistics or transfer learning (language).

4.2 Visual and Language Understanding

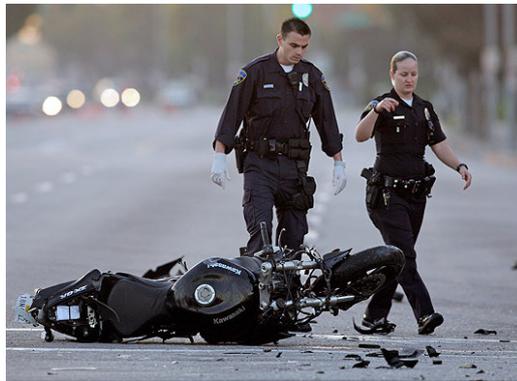
Marie-Francine Moens (KU Leuven, BE), Stephen Clark (Google DeepMind – London, GB), Luisa Coheur (INESC-ID – Lisbon, PT), Erkut Erdem (Hacettepe University – Ankara, TR), Anette Frank (Universität Heidelberg, DE), Jean Oh (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 3.0 Unported license

© Marie-Francine Moens, Luisa Coheur, Stephen Clark, Erkut Erdem, Anette Frank, Jean Oh

This working group focused on the broad topic of visual and language understanding and went deeper into the need and potential of modeling alignment between multimodal representations and their components.

An image is worth more than a thousand words: it is possible to generate many different descriptions of an image ranging from what is actually seen in the image to what could be inferred from the image. For instance, the picture of an accident (as in Figure 1) could generate the description of the accident setting, but it is more difficult to determine what the causes of the accident are or whether someone is injured.

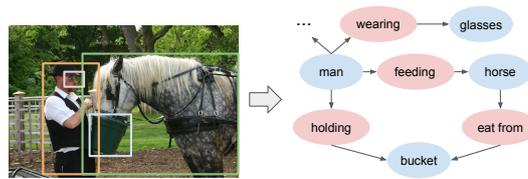


■ **Figure 1** Picture taken from [4].

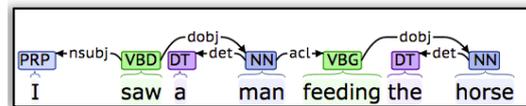
To accomplish the latter, we need some form of *reasoning* both with the *content of the image* and *prior knowledge*. Also, very often a command given to a robot is multimodal, that is, it is composed of the language command and the visual context in which the command is given. In all these cases, *knowing what information in the language is aligned with the visual data and which is not*, is valuable in understanding the command and correct action upon it.

In this working group we also discussed the properties of the visual and language data. Visual data are restricted to what could be visually perceived, but on the other hand expresses and evokes lots of knowledge. Language is often compact, ambiguous and abstract, but offers ways to be very specific, for instance, by being able to communicate negation and modal aspects of what is expressed. *Both modalities often function in a complementary way*. For instance, language leaves certain information implicit which could be learned from other modalities such as vision, acoustics or engineered knowledge (e.g, [12]). So, *complementarity helps for completing missing information*. On the other hand, overlapping or aligned information between visual and language data helps in *disambiguating* polysemous words or ambiguous image patterns.

All the above favors a multimodal representation that explicitly identifies overlaps and complementarity of the two modalities which would allow to disentangle information from the



■ **Figure 2** Scene graph from [9].



■ **Figure 3** Dependency parse (Stanford CoreNLP).

two modalities and encourage the interpretability of the representation. Such a representation would help both the semantic parsing of an image or a text as well as the generation of each modality.

The structure of a scene image or a sentence can be modeled as a graph. Scene graphs (an example shown in Figure 2) are currently very popular to capture the content of an image in terms of objects and their relationships [9, 8].

They are generated to describe an image (e.g., [9]), or images are generated from scene graphs (e.g., [3]). Graphs are also popular to structure language utterances and are the results of a (neural) dependency parse or a (neural) semantic parse (Figure 3).

A dependency parse of a sentence can in its turn easily be translated into a scene graph. As a result, these graph structures offer anchors for finding alignments between language structures and visual structures as well as between their composing components.

Humans are very good at making sense of scenes or utterances composed of objects that they have never seen before in that combination due to their ability and understanding of compositionality. A necessary condition for compositional interpretation is to recognize the components that make up a scene or language utterance, and understanding their relations. It would be interesting to *take advantage of the inherent compositionality of both images and language and integrate these properties into the learned representations*. This would entail that we can decompose representations of a whole image or video, of complete language utterances or discourses, and thus complete multimodal inputs into representations of their components.

Consequently, it would be interesting to construct structured multimodal representations. Initial attempts in this direction were made by [6], [5], [10], [7], [11]. This could lead to possible advances in *alignment, attention models, compositionality, and incremental learning*. For instance, multimodal alignment is then seen as finding relationships and correspondences between sub-components of instances from two or more modalities as in [1].

Visual data and language operate sometimes at different levels of abstraction. Possible advances in hierarchical alignment could help in finding corresponding and complementary content across modalities. An example in that direction is a hierarchical multimodal attention-based NN for image captioning as proposed by [2].

To conclude we aim at aligned, (de)composable image and linguistic representations. Explicit alignments allow to identify *overlapping vs. complementary content* and facilitate *system interpretability*. In addition, these would allow for easy and interpretable integration of symbolic (e.g., available from a knowledge resource) and continuous representations. To

reach that goal we will need more research on how to make structure more explicit through joint learning of entities and relationships in both modalities, and to align them – possibly at different levels of granularity. In the long term this could lead to inducing task-specific multimodal semantic grammars.

References

- 1 T. Baltrušaitis, C. Ahuja, and L. Morency. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, Feb 2019.
- 2 Yong Cheng, Fei Huang, Lian Zhou, Cheng Jin, Yuejie Zhang, and Tao Zhang. A hierarchical multimodal attention-based neural network for image captioning. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '17*, pages 889–892, New York, NY, USA, 2017. ACM.
- 3 Justin Johnson, Agrim Gupta, and Li Fei-Fei. Image generation from scene graphs. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 1219–1228, 2018.
- 4 Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende. Generating natural questions about an image. In *ACL. The Association for Computer Linguistics*, 2016.
- 5 Yingwei Pan, Tao Mei, Ting Yao, Houqiang Li, and Yong Rui. Jointly modeling embedding and translation to bridge video and language. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- 6 Richard Socher, Andrej Karpathy, Quoc V. Le, Christopher D. Manning, and Andrew Y. Ng. Grounded compositional semantics for finding and describing images with sentences. *Transactions of the Association for Computational Linguistics*, 2:207–218, 2014.
- 7 Ivan Vendrov, Ryan Kiros, Sanja Fidler, and Raquel Urtasun. Order-embeddings of images and language. *CoRR*, abs/1511.06361, 2016.
- 8 Yu-Siang Wang, Chenxi Liu, Xiaohui Zeng, and Alan Yuille. Scene graph parsing as dependency parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 397–407, 2018.
- 9 Danfei Xu, Yuke Zhu, Christopher B. Choy, and Li Fei-Fei. Scene graph generation by iterative message passing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- 10 Hongteng Xu, Licheng Yu, Dixin Luo, Hongyuan Zha, and Yi Xu. Dictionary learning with mutually reinforcing group-graph structures. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15*, pages 3101–3107. AAAI Press, 2015.
- 11 Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78, 2014.
- 12 Guillem Collell, Luc Van Gool, Marie-Francine Moens. Acquiring common sense spatial knowledge through implicit spatial templates. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI'18*, pages 6765–6772. AAAI Press, 2018.

4.3 Challenges in Language Generation Applications

Lucia Specia (Imperial College London, GB), Loïc Barrault (Université du Mans, FR), Thales Bertaglia (University of Sheffield, GB), Erkut Erdem (Hacettepe University – Ankara, TR), Lisa Anne Hendricks (University of California – Berkeley, US), Pecina Pavel (Charles University – Prague, CZ), Florian Metzger (Carnegie Mellon University – Pittsburgh, US), Jean Oh (Carnegie Mellon University – Pittsburgh, US)

License © Creative Commons BY 3.0 Unported license
© Lucia Specia, Loïc Barrault, Thales Bertaglia, Erkut Erdem, Lisa Anne Hendricks, Pecina Pavel, Florian Metzger, Jean Oh

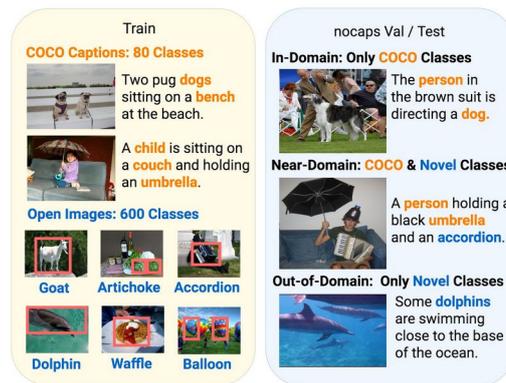
This group discussed limitations in current work in vision and language where the final objective is to generate language, given as input images/video and optionally language (e.g. the text in the source language in machine translation). The core points discussed were the types of visual information, the modality fusion architectures/approaches, how to benefit from the temporal aspect in videos, how to evaluate outputs, and a good task/dataset would be for the area. The discussion items in these topics are summarised below.

Visual information and representations. This part of the discussion focused on current work going beyond dense visual features (i.e. CNN layers) to exploit detections or proposals, mostly for objects, for example Neural Baby Talk [10] and the Bottom-Up Top-Down approach [2] for image captioning. It was suggested that using detections for places, actions, “stuff” and potentially others visual elements could be beneficial, but that there certainly are not detectors for everything. Another factor that is disregarded in current work is relevance of content in images for a given task. For example, for image description, the objects detected may not be the most important/interesting parts of the image to describe. Datasets with saliency information such as the Saliency Benchmark¹ could be useful. There is some work in this area which attempts to focus on interesting parts of the image using a loss function to make descriptions across images different from one another (e.g. [11]), or learning from eye-tracking data on how humans do it (e.g. [14]). Further structural information beyond what is in the image and its relevance for a task could also be beneficial, such as relationships between elements in the image. For texts, this can be done via dependency parsing or semantic role labelling. For images, one can consider scene graphs, such as in the Visual Genome dataset [15]. Examples from the Visual Genome dataset (<https://visualgenome.org/>) are given in Figures 4 and 5. However, this type of information is harder to generate automatically and reliability from images. A promising direction is to move beyond structure to external information, e.g. metadata or the use of multiple images for each instance. An example of work in this direction is [3], for personalized textual output. The variety of visual elements that can be detected and realistic differences in distributions between training and test sets is also important. We discussed image domain captioning via **nocaps**, a benchmark for novel object captioning at scale [1]² where 166K captions are generated for 15K images from Open Images.

Fusion Approaches/architectures. The ways in which visual information is currently fused with textual information is suboptimal. Few attempts to improve on that have been made, for example, as mentioned before, recent image captioning approaches such as Neural Baby Talk [10] and the Bottom-Up Top-Down approach [2] using object information, the use

¹ <http://saliency.mit.edu/home.html>

² <https://nocaps.org>



■ **Figure 6** The nocaps benchmark for novel object captioning (at scale) (<https://nocaps.org>).

matching at a more semantic level. For image captioning, Word Mover’s Distance, which uses word embedding as the unit to match, works better than other existing metrics [5]. For visual dialogue and storytelling, evaluation tends to be done for utterances independently. This is suboptimal as there many alternative good stories that can be generated. Text coherence models could be helpful but have not yet been investigated for this purpose.

Task/dataset. The discussion on what could be an interesting new task and dataset for language generation from images was not conclusive but some thoughts included the idea of a task that expresses common sense/creativity, and that would leverage multiple videos/images, such as multi-video textual summarization, where intermediate tasks involve measuring differences and similarities between videos. Visual dialogue with a plan/task/end-goal in mind is also an interesting direction. One could base the data collection on existing text-only dialogue datasets, e.g. [9] and [6].

To conclude, for language generation tasks, the field seems to be moving towards representing images in more structured ways, but there are a number of open questions, for example: do we need the same structure for all vision and language tasks? Does it matter if it is a tree or graph? Should it be hierarchical? Do we care about all elements and relationships? Or only salient unusual ones? Each of these open questions should lead to interesting research in the area.

References

- 1 Harsh Agrawal, Karan Desai, Xinlei Chen, Rishabh Jain, Dhruv Batra, Devi Parikh, Stefan Lee, and Peter Anderson. nocaps: novel object captioning at scale. *arXiv preprint arXiv:1812.08658*, 2018.
- 2 Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. Bottom-up and top-down attention for image captioning and visual question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6077–6086, 2018.
- 3 Cesc Chunseong Park, Byeongchang Kim, and Gunhee Kim. Attend to you: Personalized image captioning with context sequence memory networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 895–903, 2017.
- 4 Aditya Deshpande, Jyoti Aneja, Liwei Wang, Alexander Schwing, and David A Forsyth. Diverse and controllable image captioning with part-of-speech guidance. *arXiv preprint arXiv:1805.12589*, 2018.

- 5 Mert Kilickaya, Aykut Erdem, Nazli Ikizler-Cinbis and Erkut Erdem. Re-evaluating automatic metrics for image captioning. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 199–209, 2017.
- 6 He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1766–1776, 2017.
- 7 Lisa Anne Hendricks, Oliver Wang, Eli Shechtman, Josef Sivic, Trevor Darrell, and Bryan Russell. Localizing moments in video with temporal language. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1380–1390, 2018.
- 8 Anna Rohrbach, Lisa Anne Hendricks, Kaylee Burns, Trevor Darrell and Kate Saenko. Object hallucination in image captioning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4035–4045, 2018.
- 9 Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. Deal or no deal? end-to-end learning of negotiation dialogues. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2443–2453, 2017.
- 10 Jiasen Lu, Jianwei Yang, Dhruv Batra, and Devi Parikh. Neural baby talk. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7219–7228, 2018.
- 11 Zelun Luo, Boya Peng, De-An Huang, Alexandre Alahi, and Li Fei-Fei. Unsupervised learning of long-term motion dynamics for videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2203–2212, 2017.
- 12 Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- 13 Golan Pundak, Tara N Sainath, Rohit Prabhavalkar, Anjuli Kannan, and Ding Zhao. Deep context: end-to-end contextual speech recognition. *arXiv preprint arXiv:1808.02480*, 2018.
- 14 Ramprasaath R Selvaraju, Stefan Lee, Yilin Shen, Hongxia Jin, Dhruv Batra, and Devi Parikh. Taking a hint: Leveraging explanations to make vision and language models more grounded. *arXiv preprint arXiv:1902.03751*, 2019.
- 15 Ting Yao, Yingwei Pan, Yehao Li, and Tao Mei. Exploring visual relationship for image captioning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 684–699, 2018.
- 16 Shou-I Yu, Lu Jiang, and Alexander Hauptmann. Instructional videos for unsupervised harvesting and learning of action examples. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 825–828. ACM, 2014.

4.4 Modeling Human Learning

Raffaella Bernardi (University of Trento, IT), Zeynep Akata (University of Amsterdam, NL), Andrei Barbu (MIT – Cambridge, US), Ozan Caglayan (Université du Mans, FR), Stephen Clark (Google DeepMind – London, GB), Guillem Collell (KU Leuven, BE), Desmond Elliott (University of Copenhagen, DK), Raquel Fernandez (University of Amsterdam, NL), Orhan Firat (Google Inc. – Mountain View, US), Stella Frank (University of Edinburgh, GB), Frank Keller (University of Edinburgh, GB), Douwe Kiela (Facebook – New York, US), Pecina Pavel (Charles University – Prague, CZ), David Vernon (CMU Africa – Kigali, RW), Josiah Wang (University of Sheffield, GB)

License © Creative Commons BY 3.0 Unported license

© Raffaella Bernardi, Zeynep Akata, Andrei Barbu, Ozan Caglayan, Stephen Clark, Guillem Collell, Desmond Elliott, Raquel Fernandez, Orhan Firat, Stella Frank, Frank Keller, Douwe Kiela, Pecina Pavel, David Vernon, Josiah Wang

The group has focused on the connection between what is known from Cognitive Neuroscience about human learning and how such findings can be of inspiration when developing machine learning systems. We first shared our knowledge on the topic and then highlighted those research questions that would be worth to address in order to model human learning.

Human development. From Cognitive Neurosciences it has been highlighted the distinction between “developmental” and “learning” phases children go through. During the developmental phase, progresses are divided into clear not overlapping steps which are experienced by all children in the very same order (e.g., understanding before speaking). In neonates, development is manifested by the emergence of new forms of action and the acquisition of predictive control of these actions. It has been highlighted that mastery of action relies critically on prospection, i.e., the perception and knowledge of upcoming events. Repetitive practice of new actions is not focused on establishing fixed patterns of movement but on establishing the possibilities for prospective control in the context of these actions [12]. Neonats go through a “perceptual narrowing” mechanism which has been proved to be involved in language as well as in face recognition [4]. After age 2 till 10 a “synaptic pruning” process starts: 50% of the synapses are eliminated thereby increasing the efficiency of the neural network [7]. Furthermore, it has been noted that in human development a major role is played by “core abilities” which enable infants to acquire core knowledge which act as building blocks for scaffolding new cognitive abilities and more complex cognitive tasks. This core knowledge relates to perception of objects, numerosity and people.

The role of language. Besides the general overview above we zoom into the question of which is the role of language in human development. We mention the work by [8] showing that language may accelerate learning of other skills, in particular it could serve as effective priors, facilitating perception and integration of sensory information. The importance of internal talk (a.k.a., inner speech or sub-vocalization), and the findings about tongue anesthesia disrupts performance in mental tasks found in children but not in adults (after inner speech appeared) [10]. Finally, it has been pointed out the important role of language in structuring our memories has claimed in well established theories (e.g., [11]).

Continual machine learning. The importance of bringing into machine learning the developmental approach has been strongly advocated in [9]. We tried to get a picture of the advancement in ML which could be connected with the discussion summarized above on human development. In particular, the following works have been mentioned: continuation methods, viz., start with simple objectives and a simple model, then increase the task

complexity and use more neurons [3]; few-shot meta-learning, viz., start with a subset of classes, then introduce new classes, and at end all the classes [6]; knowledge distillation, viz., pre-train the model on core skills and then plug it to transfer the knowledge for learning other skills [2] or start with a big model, then compress it and specialize it on a fine-grained distinction by fine-tuning; reinforcement learning and curiosity-driven learning [5, 1].

The role of language in learning multisensory skills. What is the role of language in learning multisensory skills? Does language accelerate or amplify learning? Is language necessary to learn some specific concepts or is vision enough? Does language bring advantages to development/learning?

Relation among tasks in continual learning. Can we capture the formal relation holding between tasks useful in multimodal and multi-task settings? What should be the formal relationship between the tasks? What is a good multimodal multi-task setting? Combine pre-training with multi-task learning, how does this relate to continuation methods? Does Bayesian learning/inference have anything to offer in this context?

References

- 1 Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A. Efros. Large-scale study of curiosity-driven learning. In *ICLR*, 2019.
- 2 Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. In *NIPS*, 2015.
- 3 Caglar Gulcehre, Marcin Moczulski, Francesco Visin, and Yoshua Bengio. Mollifying networks. In *ICLR*, 2017.
- 4 Olivier Pascalis, Michelle de Haan, and Charles A. Nelson. Is face processing species-specific during the first year of life? *Science*, 296, 2002.
- 5 Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning (ICML)*, volume 2017, 2017.
- 6 Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel. Meta-learning for semi-supervised few-shot classification. In *ICLR*, 2018.
- 7 Edalmarys Santos and Chad A. Noggle. *Encyclopedia of Child Behavior and Development*, chapter Synaptic Pruning, pages 1464–1465. Springer US, 2011.
- 8 Irina Simanova, Jolien C. Francken, Floris P. de Lange, and Harold Bekkering. Linguistic priors shape categorical perception. In *Language, Cognition and Neuroscience*, volume 31, pages 159–165, 2016.
- 9 L. B. Smith and L.K. Slone. A developmental approach to machine learning. *Frontiers in Psychology*, 2017.
- 10 A. Sokolov. *Inner speech and thought*, chapter V. Springer Science and Business Media, 2012.
- 11 E. Tulving and D.M. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5):352–373, 1973.
- 12 David Vernon, Claes von Hofsten, and Luciano Fadiga. *A Roadmap for Cognitive Development in Humanoid Robots*. Springer, 2010.

4.5 Explainability

Tinne Tuytelaars (KU Leuven, BE), Luisa Coheur (INESC-ID – Lisbon, PT), Vera Demberg (Universität des Saarlandes, DE), Lisa Anne Hendricks (University of California – Berkeley, US), Dietrich Klakow (Universität des Saarlandes, DE), Jindrich Libovický (Charles University – Prague, CZ), Pranava Madhyastha (Imperial College London – GB), Marie-Francine Moens (KU Leuven, BE), Siddharth Narayanaswamy (University of Oxford, GB), Bernt Schiele (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license

© Tinne Tuytelaars, Luisa Coheur, Vera Demberg, Lisa Anne Hendricks, Dietrich Klakow, Jindrich Libovický, Pranava Madhyastha, Marie-Francine Moens, Siddharth Narayanaswamy, Bernt Schiele

The demand for and research into explainable AI is increasing. Talking about open challenges in vision and language, generating explanations of decisions and models was one that we all scored high on the research agenda. On the one hand, explanations can help us to better understand the models we are using, which may lead to more insight as well as better models per se. On the other hand, explainability is considered a crucial step towards building trust, and as such necessary to make the technology accepted in real world applications. Explanations can take many forms. Typical examples are causal graphs, graphical models, textual explanations, visual heatmaps, visualizations of CNN filters, or diagrams. They should be geared towards humans, thus intuitive and easy to interpret.

Since this is a relatively young research topic, there is some discussion about terminology, and in particular the difference between interpretation and explanation. Some consider interpretation to refer to introspection, explaining a model as a whole, whereas explanation relates to predictions or decisions for a specific input. Others consider interpretation to be more related to internal representations and understanding, whereas explanations are the formal way to communicate about this with others. For instance, a large decision tree may be interpretable, but if it's too big it's not explainable (i.e., explanation should be given at a higher level).

Explainable AI can, in general, be achieved in two ways: i) as a posthoc analysis, given a model – e.g., by discovering emerging structure, or using a surrogate model to explain the original one; or ii) by designing the models with interpretability in mind from the start – e.g., by adding an objective during training for interpretability, or by designing novel machine learning architectures that are better explainable. Some models lend themselves better for explainability than others by nature, e.g., graphical models. But explainability probably is also influenced by the data used to train a model.

A critical aspect when it comes to studying explainability is the evaluation of different methods. One criterion could be the time it takes a user to understand the interpretation. Another one may be related to how comprehensive it is, i.e., how much relevant information is retained. Since explanations are geared towards humans, it makes sense to say that humans need to be in the loop for evaluations. However, can humans really evaluate the quality of explanation? For instance, they may not like biases, even though it is a correct explanation of the model. They may also prefer explanations that are in line with their own interpretation. It's unclear how to deal with this subjectivity. The kind of explanations people like most are not necessarily the ones that are objectively the most correct. For instance, it has been shown that people prefer explanations for why a specific advertisement was selected that are not too specific.

An interesting question that popped up was, whether explanations can be used to improve the model or system itself. This is similar to the notion of reflective learning observed in humans. The fact that decisions can be explained in a consistent manner could increase the

confidence of the system in its decisions. Examples in this direction include self-labeling based on clustering, or adaptive gradient descent. When using textual explanations for visual problems or v.v., the addition of an extra modality can also improve robustness and encourage more disentangled representations.

4.6 Tasks: Creating Simulated Worlds from Existing Media

David Hogg (University of Leeds, GB), Raffaella Bernardi (University of Trento, IT), Desmond Elliott (University of Copenhagen, DK), Raquel Fernandez (University of Amsterdam, NL), Stella Frank (University of Edinburgh, GB), Marius Leordeanu (University Politehnica of Bucharest, RO), Jean Oh (Carnegie Mellon University – Pittsburgh, US), Pavel Pecina (Charles University – Prague, CZ), Lucia Specia (Imperial College London, GB), Jakob Verbeek (INRIA – Grenoble, FR), David Vernon (CMU Africa – Kigali, RW)

License © Creative Commons BY 3.0 Unported license
 © David Hogg, Raffaella Bernardi, Desmond Elliott, Raquel Fernandez, Stella Frank, Marius Leordeanu, Jean Oh, Pavel Pecina, Lucia Specia, Jakob Verbeek, David Vernon

Our discussion focused on the use of the vast repository of on-line audio-visual media, such as TV shows and movies, to create a generative model for virtual worlds. These worlds would:

- Be realistic in visual and auditory modalities;
- Follow a narrative involving simulated people (agents) behaving in a natural way;
- Be configurable for selected situations, environments, cultural and emotional norms, mirroring the content in the source media;
- Be interactive, enabling active and natural participation of real people using VR equipment.

Other potential media sources include: proceedings of world governments, customer service interactions, video-conference recordings, and AV sensors on domestic robots and autonomous vehicles.

Societal impact of such technology:

- Entertainment – interactive TV, enabling role-playing in shows; narrative transfer into new contexts (‘Romeo and Juliet’ into the ‘Friends’ genre); researching movie locations.
- Education and training – language learning; skills coaching, including generic skills; learning maths, presentation skills; robot learning.
- General Media – conceptual search/comparison on narratives, scene contexts; more empathetic agents; promotion of cultural understanding; mapping media into new cultural contexts; generated worlds as a novel communication medium between humans.
- Health – therapy for people who have difficulty recognising social cues (e.g., in ASD); culturally sensitive telemedicine.

The technology could be a transformative tool for the behavioural sciences, for example in studying human adaptability to low-fidelity agents (non-human speech patterns, prosody-only sound profiles, language abilities, masked facial appearance and expression, response latency, gender neutral agents); cultural modes of communication (distance apart, eye contact); and use of mental imagery and language in conceptual reasoning (e.g., planning and prediction of future actions).

Research challenges:

There are central challenges in achieving photo and audio realism, capturing behavioural characteristics and understanding narrative within a generative model acquired from on-line media. Beyond this, there are many interesting and fundamental research questions, such as: (1) How to create, control and manipulate virtual worlds using language cues, for increasingly complex environments; (2) How to ensure that language use acquired in the virtual world (e.g., in learning a foreign language, training a robot) is deployable in the real world, (3) The extent to which people can learn new skills from observation and non-physical interaction alone (e.g., car maintenance); (4) What is a good meaning representation of a person in learning through interaction with skilled agents; (5) How the same situation can be interpreted differently by different people, and how different views can converge; (6) What are good feedback/interaction strategies; (7) 'Teaching' versus 'doing' as pedagogical strategies; (8) What representations are needed to discover the social roles of agents, and generate dialogue to further an agent's social and long-term collective goals; (9) Is learning in a virtual environment better than learning from people in the real world, perhaps because agents are potentially friendly, better teachers, and always available, or is enabling people to avoid human contact problematic in the long term?

4.7 Tasks and Datasets for Vision and Language

Stephen Clark (Google DeepMind – London, GB), Zeynep Akata (University of Amsterdam, NL), Andrei Barbu (MIT – Cambridge, US), Loïc Barrault (Université du Mans, FR), Raffaella Bernardi (University of Trento, IT), Ozan Caglayan (Université du Mans, FR), Aykut Erdem (Hacettepe University – Ankara, TR), Erkut Erdem (Hacettepe University – Ankara, TR), Orhan Firat (Google Inc. – Mountain View, US), Anette Frank (Universität Heidelberg, DE), Stella Frank (University of Edinburgh, GB), David Hogg (University of Leeds, GB), Frank Keller (University of Edinburgh, GB), Douwe Kiela (Facebook – New York, US), Chiraag Lala (University of Sheffield, GB), Marius Leordeanu (University Politehnica of Bucharest, RO), Florian Metze (Carnegie Mellon University – Pittsburgh, US), Lucia Specia (Imperial College London, GB)

License © Creative Commons BY 3.0 Unported license

© Stephen Clark, Zeynep Akata, Andrei Barbu, Loïc Barrault, Raffaella Bernardi, Ozan Caglayan, Aykut Erdem, Erkut Erdem, Orhan Firat, Anette Frank, Stella Frank, David Hogg, Frank Keller, Douwe Kiela, Chiraag Lala, Marius Leordeanu, Florian Metze, Lucia Specia

A discussion group was led on tasks and datasets located at the intersection of Vision and Language. First of all, everyone seemed to agree that a lack of suitable datasets and tasks was a bottleneck for the development of AI systems operating at this intersection, and that developing better tasks should be a priority.

One discussion centered around the question of whether we could find “One Task to Rule Them All”, much as automatic speech recognition (ASR) seems to have found such a task in the guise of minimizing word error rate (WER). Minimizing WER is only a proxy for many of the actual tasks that we want to do with ASR, but it seems to have been a good enough proxy that it has led to clear advances in the field.

One suggestion for Vision and Language was conditional language modeling, and in particular caption generation. The point was made that we should be clear what we mean by caption generation, since this can refer to at least three distinct tasks: captions for newspaper

photographs, captions for videos, and image descriptions. Typically researchers mean the third option – an image description task – when talking about caption generation, so we focused on that.

The image description task is problematic for a number of reasons, but perhaps most fundamentally because it is often unclear which aspect of the image to focus on when providing a description, and this makes evaluation especially difficult. This relates to a broader problem of there often being a lack of a clear goal, or application, when eliciting the captions (either from a human or a machine). So one suggestion for creating better tasks and datasets is to focus on an application first, for example caption generation for visually-impaired people. It was also suggested that Visual Question Answering may be better in this regard, since it relates to an information need that a viewer of the image may have.

We then moved onto a discussion about whether it would be possible to analyse a dataset and corresponding task along a number of core dimensions, which could then be used to provide a useful summary to potential users of that dataset, and could also act as a useful guide when creating it. The idea was that there would be a relatively small number of such dimensions – perhaps as few as three – that would capture the “essence” of any dataset. However, when attempting to come up with these core dimensions, we quickly discovered that there are many important dimensions along which a dataset can vary. Examples include: number of modalities, temporal complexity, interactivity, number of agents, world complexity, linguistic complexity, existence of biases, dataset size, whether the task requires reasoning, whether the dataset tests for generalization capabilities, whether it exercises the cognitive core (intuitive physics, intuitive psychology, semantic memory).

We also attempted to analyse an existing dataset – the How2 dataset - along these dimensions. How2 is a multimodal collection of instructional videos which a number of the working group were familiar with. Again, it was surprisingly difficult to categorize this dataset along the chosen dimensions, but it was felt that this could still be a useful exercise, and that a final set of dimensions – obtained after a few cycles of use – could be a useful resource. It was also suggested that this classification may make a useful university class exercise.

Finally, we took a broader perspective on vision and language tasks, and considered what the “ultimate application” might be in this space, including the possibility of other modalities. We fixed on an “embodied Alexa”, something like the ultimate robot butler, which could tidy rooms, fix up lunch and dinner, take the children to school and so on. It was felt that computer vision technology was a long way from being in a strong enough state to be usable in such an application, but that interestingly NLP technology – especially as far as semantic parsing is concerned – might be in a better state; however, it was acknowledged that the general problem of knowledge acquisition would still need to be solved, and we’re a long way from that.

Will we have such an application in 25 years? Who knows, but the general consensus seemed to be probably not, at least in the complete form described above, but that there may be a limited, caricature of such a robot butler available (much like Alexa is currently a caricature of a fully-functioning natural language understanding system).

Participants

- Zeynep Akata
University of Amsterdam, NL
- Andrei Barbu
MIT – Cambridge, US
- Loïc Barrault
Université du Mans, FR
- Raffaella Bernardi
University of Trento, IT
- Thales Bertaglia
University of Sheffield, GB
- Ozan Caglayan
Université du Mans, FR
- Stephen Clark
Google DeepMind – London, GB
- Luísa Coheur
INESC-ID – Lisbon, PT
- Guillem Collell
KU Leuven, BE
- Vera Demberg
Universität des Saarlandes, DE
- Desmond Elliott
University of Copenhagen, DK
- Aykut Erdem
Hacettepe University –
Ankara, TR
- Erkut Erdem
Hacettepe University –
Ankara, TR
- Raquel Fernández
University of Amsterdam, NL
- Orhan Firat
Google Inc. –
Mountain View, US
- Anette Frank
Universität Heidelberg, DE
- Stella Frank
University of Edinburgh, GB
- Lisa Anne Hendricks
University of California –
Berkeley, US
- David C. Hogg
University of Leeds, GB
- Frank Keller
University of Edinburgh, GB
- Douwe Kiela
Facebook – New York, US
- Dietrich Klakow
Universität des Saarlandes, DE
- Chiraag Lala
University of Sheffield, GB
- Marius Leordeanu
University Politehnica of
Bucharest, RO
- Jindrich Libovický
Charles University – Prague, CZ
- Pranava Madhyastha
Imperial College London – GB
- Florian Metze
Carnegie Mellon University –
Pittsburgh, US
- Marie-Francine Moens
KU Leuven, BE
- Siddharth Narayanaswamy
University of Oxford, GB
- Jean Oh
Carnegie Mellon University –
Pittsburgh, US
- Pavel Pecina
Charles University – Prague, CZ
- Bernt Schiele
MPI für Informatik –
Saarbrücken, DE
- Carina Silberer
UPF – Barcelona, ES
- Lucia Specia
Imperial College London, GB
- Tinne Tuytelaars
KU Leuven, BE
- Jakob Verbeek
INRIA – Grenoble, FR
- David Vernon
CMU Africa – Kigali, RW
- Josiah Wang
University of Sheffield, GB



Logics for Dependence and Independence

Edited by

Erich Grädel¹, Phokion G. Kolaitis², Juha Kontinen³, and Heribert Vollmer⁴

- 1 RWTH Aachen, DE, graedel@informatik.rwth-aachen.de
- 2 University of California – Santa Cruz, US, kolaitis@cs.ucsc.edu
- 3 University of Helsinki, FI, juha.kontinen@helsinki.fi
- 4 Leibniz Universität Hannover, DE, vollmer@thi.uni-hannover.de

Abstract

This report documents the programme and outcomes of Dagstuhl Seminar 19031 “Logics for Dependence and Independence”. This seminar served as a follow-up seminar to the highly successful seminars “Dependence Logic: Theory and Applications” (13071) and “Logics for Dependence and Independence” (15261). A key objective of the seminar was to bring together researchers working in dependence logic and in the application areas so that they can communicate state-of-the-art advances and embark on a systematic interaction. The goal was especially to reach those researchers who have recently started working in this thriving area as well as researchers working on several aspects of database theory, separation logic, and logics of uncertainty.

Seminar January 13–18, 2019 – <http://www.dagstuhl.de/19031>

2012 ACM Subject Classification Theory of computation → Logic

Keywords and phrases dependence logic, mathematical logic, computational complexity, finite model theory, game theory

Digital Object Identifier 10.4230/DagRep.9.1.28

Edited in cooperation with Katrin M. Dannert

1 Executive Summary

Erich Grädel (RWTH Aachen, DE)

Phokion G. Kolaitis (University of California – Santa Cruz, US)

Juha Kontinen (University of Helsinki, FI)

Heribert Vollmer (Leibniz Universität Hannover, DE)

License  Creative Commons BY 3.0 Unported license
© Erich Grädel, Phokion G. Kolaitis, Juha Kontinen and Heribert Vollmer

Brief Introduction to the Topic

Dependence and independence are interdisciplinary notions that are pervasive in many areas of science. They appear in domains such as mathematics, computer science, statistics, quantum physics, and game theory. The development of logical and semantical structures for these notions provides an opportunity for a systematic approach, which can expose surprising connections between different areas, and may lead to useful general results.

Dependence Logic is a tool for modeling dependencies and interaction in dynamical scenarios. Reflecting this, it has higher expressive power and complexity than classical logics used for these purposes previously. Algorithmically, first-order dependence logic corresponds exactly to the complexity class NP and to the so-called existential fragment of second-order logic. Since the introduction of dependence logic in 2007, the framework has been



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

Logics for Dependence and Independence, *Dagstuhl Reports*, Vol. 9, Issue 01, pp. 28–46

Editors: Erich Grädel, Phokion G. Kolaitis, Juha Kontinen, and Heribert Vollmer



Dagstuhl Reports

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

generalized, e. g., to the contexts of modal, intuitionistic, and probabilistic logic. Moreover, interesting connections have been found to complexity theory, database theory, statistics, and dependence logic has been applied in areas such as linguistics, social choice theory, and physics. Although significant progress has been made in understanding the computational side of these formalisms, still many central questions remain unsolved so far. In addition to addressing the open questions, the seminar also aimed at boosting the exchange of ideas and techniques between dependence logic and its application areas.

Organization of the Seminar and Activities

The workshop brought together 40 researchers from mathematics, database theory, natural language semantics, and theoretical computer science. The participants consisted of both senior and junior researchers, including a number of postdocs and advanced graduate students.

Participants were invited to present their work and to communicate state-of-the-art advances. Over the five days of the workshop, 27 talks of various lengths took place. Introductory and tutorial talks of 90-60 minutes were scheduled prior to the workshop. Most of the remaining slots were filled, mostly with shorter talks, as the workshop commenced. The seminar ended with an open problems and perspectives session. The organizers considered it important to leave ample free time for discussion.

The tutorial talks were scheduled during the beginning of the week in order to establish a common background for the different communities that came together for the workshop.

The presenters and topics were:

- Miika Hannula: Team semantics
- Val Tannen: Provenance
- Dan Suciu: Probabilistic databases
- Meghyn Bienvenu: Constraints in ontology based databases
- David Pym: Resource semantics
- Magdalena Ortiz: Complete and incomplete information in knowledge-enriched databases
- Jef Wijsen: Database repairs

In addition, the seminar consisted of 20 shorter contributed talks, addressing various topics concerning expressibility, axiomatizability, complexity and applications of team-based logics.

The last session of the workshop was devoted to open problems and consisted of contributions by Phokion Kolaitis, Jouko Väänänen and Juha Kontinen presenting questions about decidability and axiomatizability of the implication problem of various fragments of dependence and independence logic, Joachim Biskup addressing decidable first-order prefix classes in the database context, Heribert Vollmer presenting open relationships among various counting classes related to team-based logics, Lauri Hella talking about union-closed properties in Σ_1^1 , and finally Raine Rönnholm addressing relationships between fragments of inclusion logic and greatest fixed-point logic.

The workshop ended with a discussion of future perspectives of the study of logics for dependence and independence.

The workshop achieved its aim of bringing together researchers from various related communities to share state-of-the-art research. The organizers left ample time outside of this schedule of talks and many fruitful discussions between participants took place throughout the afternoons and evenings.

Concluding Remarks and Future Plans

The organizers regard the workshop as a great success. Bringing together researchers from different areas fostered valuable interactions and led to fruitful discussions. Feedback from the participants was very positive as well.

Finally, the organizers wish to express their gratitude toward the Scientific Directorate of the Center for its support of this workshop.

2 Table of Contents

Executive Summary

Erich Grädel, Phokion G. Kolaitis, Juha Kontinen and Heribert Vollmer 28

Overview of Talks

| | |
|--|----|
| Team semantics | |
| <i>Mika Hannula</i> | 33 |
| The Semiring Framework for Provenance | |
| <i>Val Tannen</i> | 33 |
| Probabilistic databases | |
| <i>Dan Suciu</i> | 34 |
| A brief introduction to ontology-mediated query answering | |
| <i>Meghyn Bienvenu</i> | 34 |
| Logic as a modelling technology: resource semantics, systems modelling, and security | |
| <i>David J. Pym</i> | 34 |
| Complete and incomplete information in knowledge-enriched databases | |
| <i>Magdalena Ortiz</i> | 35 |
| Database Repairs | |
| <i>Jef Wijsen</i> | 36 |
| Complexity Classifications of Functional Dependencies in Database Repairing | |
| <i>Benny Kimelfeld</i> | 36 |
| Semiring Provenance for Logics with Team Semantics | |
| <i>Erich Grädel</i> | 36 |
| An atom's worth of anonymity | |
| <i>Jouko Väänänen</i> | 37 |
| On the expressive power of anonymity atoms | |
| <i>Raine Rönnholm</i> | 37 |
| Various forms of independence in possibility theory: An overview | |
| <i>Henri Prade</i> | 37 |
| Initial Steps into Parametrised Complexity of Dependence Logic | |
| <i>Yasir Mahmood</i> | 38 |
| Dependency Concepts up to Equivalence | |
| <i>Matthias Hoelzel</i> | 39 |
| Variations on a Causality Theme in Data Management | |
| <i>Leopoldo Bertossi</i> | 39 |
| Logics with Multiteam Semantics | |
| <i>Richard Wilke</i> | 40 |
| Probabilistic team semantics | |
| <i>Jonni Virtema</i> | 40 |
| Propositional Union Closed Team Logics | |
| <i>Fan Yang</i> | 41 |

| | |
|---|----|
| Set-valued Dependence versus Value Independence and Inferences in Pure Flat Attribute Universes for Relational Database Schema Design | |
| <i>Joachim Biskup</i> | 42 |
| Separation Logic and Team Semantics | |
| <i>Erich Grüdel</i> | 42 |
| Temporal Logics for Hyperproperties | |
| <i>Bernd Finkbeiner</i> | 42 |
| Holistic treatment of syntax and semantics | |
| <i>Bernhard Thalheim</i> | 43 |
| Reasoning about dependence and independence in aggregation problems | |
| <i>Eric J. Pacuit</i> | 43 |
| On matrices and K-relations | |
| <i>Jan Van den Bussche</i> | 43 |
| Approximate dependency atoms | |
| <i>Åsa Hirvonen</i> | 44 |
| Counting of Teams in First-Order Team Logics | |
| <i>Juha Kontinen</i> | 44 |
| Finite-State Dependence | |
| <i>Dietmar Berwanger</i> | 44 |
| The exact status of database semantics? | |
| <i>Joachim Biskup</i> | 45 |
| Participants | 46 |

3 Overview of Talks

3.1 Team semantics

Miika Hannula (University of Helsinki, FI)

License  Creative Commons BY 3.0 Unported license
© Miika Hannula

Team semantics provides a framework for modern logics of dependence and independence. In this tutorial talk we cover the basic theory for logics in team semantics. We also give a quick survey to some of the recent trends and developments in the field.

3.2 The Semiring Framework for Provenance

Val Tannen (University of Pennsylvania – Philadelphia, US)

License  Creative Commons BY 3.0 Unported license
© Val Tannen

Data provenance: Imagine a computational process that uses a complex input consisting of multiple items. The granularity and nature of “input item” can vary significantly. It can be a single tuple, a database table, or a whole database. It can be a spreadsheet describing an experiment, a laboratory notebook entry, or another form of capturing annotation by humans in software. It can also be a file, or a storage system component. It can be a parameter used by a module in a scientific workflow. It can also be a configuration rule used in software-defined routing or in a complex network protocol. Or it can be a configuration decision made by a distributed computation scheduler (think map-reduce). Provenance analysis allows us to understand how these different input items affect the output of the computation. When done appropriately, such analysis can be further used, for example,

- A1:** to figure out how much to trust the output, assuming that we may trust some input items more than others;
- A2:** to minimize the cost of obtaining the output, assuming that one has to pay for the input items;
- A3:** to figure out the clearance level required for accessing the output, assuming that we know the clearance levels for the input items;
- A4:** to compute the probabilistic distribution of the output, assuming that we know the distributions of the input items;
- A5:** to figure out if the output might change (and therefore whether output maintenance is necessary) when certain input items change; or
- A6:** to track back and find the input items at fault, assuming the output is somehow wrong.

We shall have occasion below to refer to the applications A1-A6 just listed. In practice, a computational process will produce a collection of output items and the applications above become more interesting when we realize that we may get different analyses for each output item.

Now, observe that these applications should not rely on just a trace, or a log, of a specific execution of the computational process itself. Some approach to applications A3 and A5 could probably be devised using just execution traces, but we should still worry whether repeating the execution with the same input items but with a differently optimized execution platform will produce the same provenance. At the same time, we cannot simply peg provenance as

a kind of static analysis of the underlying program since it will also depend on the input items themselves. This dual static-dynamic nature makes provenance analysis particularly interesting, especially so for database computations.

Provenance of FOL model-checking: Given a first-order sentence, a model-checking computation tests whether the sentence holds true in a given finite structure. Data provenance extracts from this computation an abstraction of the manner in which its result depends on the data items that describe the model. Previous work on provenance was, to a large extent, restricted to the negation-free fragment of first-order logic and showed how provenance abstractions can be usefully described as elements of commutative semirings — most generally as multivariate polynomials with positive integer coefficients. We introduce a novel approach to dealing with negation and a corresponding commutative semiring of polynomials with dual indeterminates. These polynomials are used to perform reverse provenance analysis, i.e., finding models that satisfy various properties under given provenance tracking assumptions.

3.3 Probabilistic databases

Dan Suciu (University of Washington – Seattle, US)

License  Creative Commons BY 3.0 Unported license
© Dan Suciu

We examine the implication problem between soft constraints in two settings. The first uses a probability distribution on models, and is based on the work done in probabilistic databases and in Markov Logic Networks (MLN). The second is based on using information theoretic measures to quantify the degree of a constraint.

3.4 A brief introduction to ontology-mediated query answering

Meghyn Bienvenu (University of Bordeaux, FR)

License  Creative Commons BY 3.0 Unported license
© Meghyn Bienvenu

Recent years have seen an increasing interest in ontology-mediated query answering (OMQA), in which the semantic knowledge provided by an ontology is exploited when querying data. In this talk, I will give a short introduction to this area, focusing on ontologies formulated using description logics. After introducing description logics and the OMQA problem, I will provide a brief overview of the main algorithmic techniques and the complexity landscape.

3.5 Logic as a modelling technology: resource semantics, systems modelling, and security

David J. Pym (University College London, GB)

License  Creative Commons BY 3.0 Unported license
© David J. Pym

The development of BI, the logic of bunched implications, together with its resource semantics, led to the formulation of Separation Logic, which forms the basis of the Infer program analyser deployed in Facebook’s code production. However, this rather successful story sits within a

broader, quite systematic logical context. I will review the (family of) logics – including modal logics, logics for layered graphs, and process logics – that are supported by resource semantics, explaining their more-or-less uniform meta-theoretic basis and illustrating their uses in a range of modelling applications, including access control, systems security, and workflow simulation. Many references are available at: <http://www.cs.ucl.ac.uk/staff/D.Pym/>.

3.6 Complete and incomplete information in knowledge-enriched databases

Magdalena Ortiz (TU Wien, AT)

License © Creative Commons BY 3.0 Unported license
© Magdalena Ortiz

Joint work of Shqiponja Ahmetaj, Labinot Bajraktari, Nhung Ngo, Magdalena Ortiz, Mantas Simkus
Main reference Labinot Bajraktari, Magdalena Ortiz, Mantas Simkus: “Combining Rules and Ontologies into Clopen Knowledge Bases”, in Proc. of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pp. 1728–1735, AAAI Press, 2018.
URL <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16991>

Ontologies are background theories expressing domain knowledge written in a logical formalism that supports automated inference (i.e., logics with a decidable entailment problem, such as description logics). Ontologies have been successfully used for inferring better answers from incomplete data, but the usual first-order semantics used in this setting, which assumes that all data is incomplete, can sometimes be too weak and not give all expected answers. To overcome this problem, closed predicates have been considered in the description logics literature. In a nutshell, closed predicates enhance an ontology with a list of predicates that are assumed complete, analogously to master tables in databases. This talk summarizes some of the challenges that closed predicates pose, including non-monotonicity of the consequence relation and increased computational complexity of reasoning [3]. We discuss some rewritings of ontology-mediated queries with closed predicates into Datalog extensions [1], and briefly describe a very rich knowledge representation language that supports closed predicates and extends some classic hybrid languages combining Datalog and description logics [2].

References

- 1 Shqiponja Ahmetaj, Magdalena Ortiz, and Mantas Simkus. Polynomial Datalog Rewritings for Expressive Description Logics with Closed Predicates. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016*, pages 878–885. AAAI Press, 2016.
- 2 Labinot Bajraktari, Magdalena Ortiz, and Mantas Simkus. Combining Rules and Ontologies into Clopen Knowledge Bases. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 1728–1735. AAAI Press, 2018.
- 3 Nhung Ngo, Magdalena Ortiz, and Mantas Simkus. Closed predicates in description logics: Results on combined complexity. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016*, pages 237–246. AAAI Press, 2016.

3.7 Database Repairs

Jef Wijsen (University of Mons, BE)

License  Creative Commons BY 3.0 Unported license
© Jef Wijsen

Research in database repairing and consistent query answering started with the seminal paper [Arenas, Bertossi, and Chomicki, PODS 1999]. In this talk, we survey twenty years of research in this field, with a particular focus on the following topics:

- database dependencies that have appeared in logics for dependence and independence;
- a generic definition of the notion of database repair;
- the computational complexity of the problem known as symmetric-difference repair checking, for different classes of database dependencies;
- the computational complexity of symmetric-difference consistent query answering with respect to conjunctive queries and different classes of database dependencies;
- a fine-grained complexity classification for consistent query answering to self-join-free conjunctive queries with respect to key dependencies.

3.8 Complexity Classifications of Functional Dependencies in Database Repairing

Benny Kimelfeld (Technion – Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Benny Kimelfeld

The talk describes our research on the computational complexity of problems that arise in reasoning about the inconsistency of databases. To that extent, an inconsistent database is a database that violates a set of integrity constraints, and a repair is a consistent database that is obtained from the inconsistent one via a legitimate sequence of repairing operations. Focusing on functional dependencies as constraints and tuple deletions as repairing operations, I discuss several related computational problems. One problem is that of repairing through a minimal number of deletions. Another problem is that of finding a most probable repair when tuples are associated with probabilities. Other problems involve counting and enumerating set-minimal repairs, possibly in the presence of preferences among tuples. In each problem, the talk focuses on the classification of the constraint sets into ones that admit a tractable solution, and ones that are provably hard.

3.9 Semiring Provenance for Logics with Team Semantics

Erich Grädel (RWTH Aachen, DE)

License  Creative Commons BY 3.0 Unported license
© Erich Grädel

Joint work of Erich Grädel, Lukas Huwald

We extend the approach of provenance analysis by interpretations in commutative semirings to logics of dependence and independence. We investigate issues such as locality, closure properties, game based analysis, and expressive power in this wider context. It turns out that for a smooth theory the cases of idempotent or absorptive semirings seem particularly adequate.

3.10 An atom's worth of anonymity

Jouko Väänänen (University of Helsinki, FI)

License  Creative Commons BY 3.0 Unported license
© Jouko Väänänen

I observe that team semantics is very suitable for the study of k -anonymity and other concepts motivated by privacy concerns. To this end I introduce what I call k -anonymity atoms. The 2-anonymity atom has been already introduced by Galliani under the name of non-dependence atom. k -anonymity atoms and even stronger related atoms have been introduced by Grädel and Hegselmann under the name of forking atom. I give a complete axiomatization of 2-anonymity atoms and suggest an axiomatization of k -anonymity atoms. I also conjecture that there is a complete axiomatization of the anonymity atom together with the dependence atom. By results of Galliani, anonymity logic, i.e. the extension of first order logic by anonymity atoms, is equivalent to inclusion logic.

3.11 On the expressive power of anonymity atoms

Raine Rönholm (Tampere University, FI)

License  Creative Commons BY 3.0 Unported license
© Raine Rönholm

Anonymity atoms (originally called non-dependence atoms) were introduced by Pietro Galliani in 2012. Galliani has shown that when anonymity atoms are added to first order logic with team semantics, we obtain an equivalent logic with inclusion logic. The truth condition of the anonymity atom $\text{an}(x, y)$ intuitively states that the truth of the corresponding dependence atom $\text{=}(x, y)$ is “violated” for each value of x in the team. That is, for each value of x , there exist assignments s and s' which agree on x but have a different value for y . These atoms can be further generalized to so-called m -anonymity atoms which state that for each value of x there are at least m different values for y in the team. We obtain k -ary (m)-anonymity atoms by allowing $(k - 1)$ -tuples of variables in the place of the variable x in the atom $\text{an}(x, y)$. We study how the expressive power of (m)-anonymity atoms is affected by making these restrictions on the arity. We present some new results which are obtained by making comparisons with the arity fragments of inclusion logic and the relational arity fragments of existential second order logic.

3.12 Various forms of independence in possibility theory: An overview

Henri Prade (University of Toulouse, FR)

License  Creative Commons BY 3.0 Unported license
© Henri Prade

The idea of independence has been mainly discussed in two settings: probability and logic. In probability, the independence between events is a symmetrical notion, since saying that B is independent of A reads $\text{Prob}(B|A) = P(B)$ which is equivalent to $P(A \& B) = P(A)P(B)$ (& stands for conjunction). Moreover we can simplify probabilistic calculations by knowing or assuming that variables are (conditionally) independent. Logical independence differs from

stochastic independence, and the situation is not the same, since it is rather dependencies that are asserted in logic. In this presentation, we provide an overview of various notions of independence that have been defined in possibility theory, a setting appropriate for modelling epistemic uncertainty, and which also remains close to logic. First, independence between events is no longer necessarily symmetrical in possibility theory, and is useful for introducing independence information in nonmonotonic reasoning. Independence between variables is of a different nature: it may be symmetrical and generalize logical independence. The relation with possibilistic functional dependencies that have been recently introduced and shown useful in database design will be also discussed.

References

- 1 S. Benferhat, D. Dubois, H. Prade. Possibilistic independence and plausible reasoning. In *Proc. of the Conf. on Foundations and Applications of Possibility Theory (FAPT'95)*, Ghent, Belgium, 13/12/1995-15/12/1995, World Scientific, p. 47–63, 1995.
- 2 S. Benferhat, D. Dubois, H. Prade. Practical Handling of Exception-Tainted Rules and Independence Information in Possibilistic Logic. In *Applied Intelligence*, 9, 101–127, 1998.
- 3 S. Benferhat, D. Dubois, H. Prade. The possibilistic handling of irrelevance in exception-tolerant reasoning. In *Annals of Mathematics and Artificial Intelligence*, 35, 29–61, 2002.
- 4 N. Ben Amor, S. Benferhat, D. Dubois, H. Geffner and H. Prade. Independence in qualitative uncertainty frameworks. In *Proc. 7th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2000)*, Breckenridge, Co., April 11-15, 2000, 235–246.
- 5 N. Ben Amor, K. Mellouli, S. Benferhat, D. Dubois, H. Prade. A theoretical framework for possibilistic independence in a weakly ordered setting. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (2), 117–155, 2002.
- 6 D. Dubois, L. Fariñas del Cerro, A. Herzig, H. Prade. An ordinal view of independence with application to plausible reasoning. In *Proc. of the 10th Conf. on Uncertainty in Artificial Intelligence* (R. Lopez de Mantaras, D. Poole, eds.), Seattle, WA, July 29-31, 1994, 195–203.
- 7 D. Dubois, L. Fariñas, A. Herzig, H. Prade. Qualitative relevance and independence: A roadmap. In *Proc. 15th Int. Joint Conf. on A.I.*, Nagoya, 62-67. Extended version: A roadmap of qualitative independence. *Fuzzy Sets, Logics and Reasoning about Knowledge* (Dubois, D., Prade, H., Klement, E.P., eds.), Kluwer Academic Publ., 325–350, 1999.
- 8 D. Dubois, E. Pap, H. Prade. Pseudo-additive measures and the independence of events. In *Technologies for Constructing Intelligent Systems 1 Tools* (Bouchon-Meunier, B. et Gutierrez-Rios, J. et Magdalena, L. et Yager, R. R., Eds.), Physica-Verlag, 179–192, 2002.
- 9 Sebastian Link, Henri Prade. Possibilistic Functional Dependencies and Their Relationship to Possibility Theory. In *IEEE Trans. Fuzzy Systems*, 24(3), 757–763, 2016.
- 10 Sebastian Link, Henri Prade. Relational Database Schema Design for Uncertain Data. *CIKM 2016* 1211–1220.

3.13 Initial Steps into Parametrised Complexity of Dependence Logic

Yasir Mahmood (Leibniz Universität Hannover, DE)

License  Creative Commons BY 3.0 Unported license
© Yasir Mahmood

In this talk, we present work in progress. We introduce parametrised complexity for team-based logic and start with the propositional dependence logic.

- Questions addressed are: Satisfiability, Model-checking & Validity of a PDL-formula.
- Potential parameters are: Team-size, number of splits, tree-width, tree-depth and number of variables.

We present some initial results: for (MC)

1. An application of Courcelle’s theorem to model checking (for both, lax and strict semantics)
2. An FPT-algorithm when the parameter is team-size + tree-depth and team-size + #splits
3. (For SAT) when parameter is #variables.

Furthermore, we present some observations regarding the satisfiability problem and close with some questions that will be addressed in the future.

3.14 Dependency Concepts up to Equivalence

Matthias Hoelzel (RWTH Aachen, DE)

License © Creative Commons BY 3.0 Unported license
© Matthias Hoelzel

Joint work of Matthias Hoelzel, Erich Grädel

Main reference Erich Grädel, Matthias Hoelzel: “Dependency Concepts up to Equivalence”, in Proc. of the 27th EACSL Annual Conference on Computer Science Logic, CSL 2018, September 4-7, 2018, Birmingham, UK, LIPIcs, Vol. 119, pp. 25:1–25:21, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.

URL <http://dx.doi.org/10.4230/LIPIcs.CSL.2018.25>

We study logics with dependency statements that cannot distinguish elements up to equality, but only up to a given equivalence relation. We analyse the power of such logics, by identifying equally expressive fragments of existential second-order logic or greatest fixed-point logic, with relations that are closed under a given equivalence.

3.15 Variations on a Causality Theme in Data Management

Leopoldo Bertossi (Carleton University – Ottawa, CA & RelationalAI Inc., CA)

License © Creative Commons BY 3.0 Unported license
© Leopoldo Bertossi

Joint work of Leopoldo Bertossi, Babak Salimi

The presentation reviews several problems and results in relation to the specification and computation of causes for query answers in data management. In particular, the problems of computing causes, their responsibilities, and maximum-responsibility causes are considered, and results for them are obtained by exploiting a connection between DB causality and repairs of databases that violate integrity constraints (ICs). Also answer-set programs (ASPs) are proposed for the specification of causes and their responsibilities. They are based on ASPs that specify database repairs. The problems of specifying and computing causes under ICs are introduced and some results are presented. Finally, a formalization of causes for query answers at the attribute level is proposed.

References

- 1 L. Bertossi and B. Salimi. *From Causes for Database Queries to Repairs and Model-Based Diagnosis and Back*. Theory of Computing Systems, 2017, 61(1):191–232.
- 2 L. Bertossi and B. Salimi. *Causes for Query Answers from Databases: Datalog Abduction, View-Updates, and Integrity Constraints*. International Journal of Approximate Reasoning, 2017, 90:226–252.
- 3 L. Bertossi. *Characterizing and Computing Causes for Query Answers in Databases from Database Repairs and Repair Programs*. Proc. FoIKs, 2018, Springer LNCS 10833, pp. 55–76.

3.16 Logics with Multiteam Semantics

Richard Wilke (RWTH Aachen, DE)

License  Creative Commons BY 3.0 Unported license
© Richard Wilke

Joint work of Erich Grädel, Richard Wilke

Team semantics is the mathematical basis of modern logics of dependence and independence. In contrast to classical Tarski semantics, a formula is evaluated not for a single assignment of values to the free variables, but on a set of such assignments, called a team. Team semantics is appropriate for a purely logical understanding of dependency notions, where only the presence or absence of data matters, but based on sets, it does not take into account multiple occurrences of data values. It is therefore insufficient in scenarios where such multiplicities matter, in particular for reasoning about probabilities and statistical independencies. Therefore, an extension from teams to multiteams (i.e. multisets of assignments) has been proposed by several authors.

We aim at a systematic development of logics of dependence and independence based on multiteam semantics. We study atomic dependency properties of finite multiteams and discuss the appropriate meaning of logical operators for multiteam semantics, so as to extend the atomic dependencies to full-fledged logics for reasoning about dependence and independence in a multiteam setting. We compare the properties and expressive power of a number of different logics with team and multiteam semantics. It turns out that the relationship between multiteam semantics, team semantics, and classical Tarski semantics, and the study of the expressive power of logics with multiteam semantics are more delicate and more interesting than one might expect. In particular, with multiteam semantics, inclusion and exclusion logic does not correspond to independence logic.

3.17 Probabilistic team semantics

Jonni Virtema (Hasselt University, BE)

License  Creative Commons BY 3.0 Unported license
© Jonni Virtema

Joint work of Arnaud Durand, Miika Hannula, Åsa Hirvonen, Juha Kontinen, Vadim Kulikov, Arne Meier and Jonni Virtema

Main reference Arnaud Durand, Miika Hannula, Juha Kontinen, Arne Meier, Jonni Virtema: “Probabilistic Team Semantics”, in Proc. of the Foundations of Information and Knowledge Systems – 10th International Symposium, FoIKS 2018, Budapest, Hungary, May 14-18, 2018, Proceedings, Lecture Notes in Computer Science, Vol. 10833, pp. 186–206, Springer, 2018.

URL https://doi.org/10.1007/978-3-319-90050-6_11

We review recent work on probabilistic team semantics [1, 2]. Probabilistic team semantics is built compositionally upon the notion of a probabilistic team, that is, a probability distribution over variable assignments. This framework allows the study of logical and probabilistic dependencies simultaneously. Adapting probabilistic team semantics recovers some desired properties of the so-called strict team semantics. Probabilistic team semantics has also a close connection to the area of meta finite model theory; the expressive powers of related logics are captured by variants of two-sorted logics with arithmetic operations on the second numeric sort.

References

- 1 Durand, A., Hannula, M., Kontinen, J., Meier, A., Virtema, J.: Probabilistic team semantics. In: FoIKS. Lecture Notes in Computer Science, vol. 10833, pp. 186–206. Springer (2018). 10.1007/978-3-319-90050-6_11
- 2 Hannula, M., Hirvonen, Å., Kontinen, J., Kulikov, V., Virtema, J.: Facets of distribution identities in probabilistic team semantics. *CoRR*, abs/1812.05873, 2018. Accepted to the 16th edition of the European Conference on Logics in Artificial Intelligence, JELIA 2019.

3.18 Propositional Union Closed Team Logics

Fan Yang (University of Helsinki, FI)

License © Creative Commons BY 3.0 Unported license
© Fan Yang

Logics based on team semantics (also called team logics) often have interesting closure properties. For example, dependence logic is closed downwards, meaning that the truth of a formula on a team is preserved under taking subteams. In this talk, we discuss propositional team logics that are closed under unions, meaning that if two teams both satisfy a formula, then their union also satisfies the formula. Inclusion logic [1] is closed under unions. Other known union closed logics are classical logic extended with anonymity atoms (introduced very recently by Väänänen to characterize anonymity in the context of privacy), or with the relevant disjunction (introduced by Rönholm [3], and also named nonempty disjunction by some other authors [2, 5]).

While propositional downwards closed team logics are well studied (e.g., [4]), propositional union closed team logics are not understood very well yet. It follows from [2] that propositional inclusion logic (PInc) with extended inclusion atoms is expressively complete, and PInc is thus expressively equivalent to classical logic extended with relevant disjunction (PU), which is shown to be also expressively complete in [5]. We show in this talk that classical logic extended with anonymity atoms (PAm) is also expressively complete, and PInc with slightly less general inclusion atoms is already expressively complete. From the expressive completeness, we will derive the interpolation theorem for these logics. We also provide axiomatizations for these logics, which are lacking in the literature.

References

- 1 P. Galliani, *Inclusion and exclusion in team semantics: On some logics of imperfect information*. *Annals of Pure and Applied Logic* 163, 1 (January 2012), 68–84.
- 2 L. Hella, and J. Stumpf, *The expressive power of modal logic with inclusion atoms*. In *Proceedings of the 6th GandALF (2015)*, pp. 129–143.
- 3 R. Rönholm, *Arity Fragments of Logics with Team Semantics*. PhD thesis, University of Helsinki, 2018.
- 4 F. Yang, and J. Väänänen, *Propositional logics of dependence*. *Annals of Pure and Applied Logic* 167,7 (July 2016), 557–589.
- 5 F. Yang, and J. Väänänen, *Propositional team logics*. *Annals of Pure and Applied Logic* 168, 7 (July 2017), 1406–1441.

3.19 Set-valued Dependence versus Value Independence and Inferences in Pure Flat Attribute Universes for Relational Database Schema Design

Joachim Biskup (TU Dortmund, DE)

License  Creative Commons BY 3.0 Unported license

© Joachim Biskup

Joint work of Joachim Biskup, Sebastian Link

Main reference Joachim Biskup: “Inferences of Multivalued Dependencies in Fixed and Undetermined Universes”, *Theor. Comput. Sci.*, Vol. 10, pp. 93–105, 1980.

URL [https://doi.org/10.1016/0304-3975\(80\)90075-4](https://doi.org/10.1016/0304-3975(80)90075-4)

Dependence and independence are crucial notions for layered relational database schema design, which includes decisions on data formats and their consequences for storing data either together or separately. The design process considers at least two layers, first conceptual modeling with pure flat attributes and then relational formalization with flat predicates of fixed arity (a variant of FOL formatting). Reviewing the expressive means of the layers and how constraints are first specified and later converted, we observe that a set-valued dependence constraint between pure flat attributes might become closely related to a value independence constraint for a fixed predicate. Originally studied for relational database schema design guided by multivalued dependencies and later on also exploited for other intensional data formatting tasks, such relationships have to be reflected by “appropriate inference”, distinguishing between application-driven reasoning on the layer of pure flat attributes and formatting-driven reasoning on the layer of fixed flat predicates.

3.20 Separation Logic and Team Semantics

Erich Grädel (RWTH Aachen, DE)

License  Creative Commons BY 3.0 Unported license

© Erich Grädel

In this short talk, I discussed how (a particular variant of) separation logic can be understood in terms of team semantics. I hope that this may lead to a deeper study of connections between separation logic (and its cousins) with logics of dependence and independence.

3.21 Temporal Logics for Hyperproperties

Bernd Finkbeiner (Universität des Saarlandes, DE)

License  Creative Commons BY 3.0 Unported license

© Bernd Finkbeiner

A hyperproperty is a set of sets of traces. Hyperproperties are commonly used in information flow security to express requirements of the type “the public output of a system must not depend on its secret inputs” such as noninterference, noninference, or observational determinism. In this talk, I will give an overview on HyperLTL, the extension of linear-time temporal logic (LTL) to hyperproperties, and related logics such as LTL with team semantics and first-order logic over sets of traces.

3.22 Holistic treatment of syntax and semantics

Bernhard Thalheim (CAU Kiel, DE)

License © Creative Commons BY 3.0 Unported license
© Bernhard Thalheim

Database research as well as Computer Science research separate syntax and semantics in a two-step definition mould. Syntax is considered to be a “firstness” property and semantics some kind of “secondness” (in the sense of Peirce).

We propose a holistic treatment of syntax and semantics similar to the one that natural languages use.

The holistic treatment is necessary for model development and utilisation.

It can be considered as an specific form of team semantics that is extended by context.

3.23 Reasoning about dependence and independence in aggregation problems

Eric J. Pacuit (University of Maryland – College Park, US)

License © Creative Commons BY 3.0 Unported license
© Eric J. Pacuit
Joint work of Eric J. Pacuit, Fan Yang

Notions of dependence and independence are central to many key results in preference aggregation and opinion pooling (aggregating probabilistic judgements). In this talk, I will briefly discuss joint work with Fan Yang on the formalization of Arrow’s Theorem in an independence logic. My goal in this talk is to examine other results in the preference and judgement aggregation literature that are amenable to formalization in dependence and/or independence logic. I will focus on capturing notions of domain restrictions from the social choice literature and impossibility results about the preservation of independence when aggregating probabilistic judgements.

3.24 On matrices and K-relations

Jan Van den Bussche (Hasselt University, BE)

License © Creative Commons BY 3.0 Unported license
© Jan Van den Bussche
Joint work of Jan Van den Bussche, Robert Brijder, Marc Gyssens

MATLANG, proposed by Brijder, Geerts, Van den Bussche and Weerwag at ICDT 2018, is an algebra for querying matrix databases. K -relations can represent matrices as well, and a semantics for the positive relational algebra on K -relations was defined by Green, Karvounarakis and Tannen at PODS 2007. We refer to this algebra as ARA. One can easily translate MATLANG into ARA, and only relations with at most three attributes are needed for this translation; we denote this by ARA(3). We prove a converse result: every binary ARA(3) query over binary relations can be expressed in MATLANG.

3.25 Approximate dependency atoms

Åsa Hirvonen (University of Helsinki, FI)

License  Creative Commons BY 3.0 Unported license
© Åsa Hirvonen

Joint work of Åsa Hirvonen, Juha Kontinen, Arno Pauly

We give a brief overview of approximate dependency atoms in infinite teams over metric spaces. The motivation is, on one hand, to get better behaved notions of dependency when one looks at infinite metric teams from a computational point of view, on the other, to develop notions of approximate dependencies tied to “almost correct” data.

3.26 Counting of Teams in First-Order Team Logics

Juha Kontinen (University of Helsinki, FI)

License  Creative Commons BY 3.0 Unported license
© Juha Kontinen

Joint work of Juha Kontinen, Anselm Haak, Fabian Müller, Heribert Vollmer, Fan Yang

We study descriptive complexity of counting complexity classes in the range from $\#P$ to $\#\text{-NP}$. A corollary of Fagin’s characterization of NP by existential second-order logic is that $\#P$ can be logically described as the class of functions counting satisfying assignments to free relation variables in first-order formulae. In this talk we extend this study to classes beyond $\#P$ and extensions of first-order logic with team semantics. Our results show that the class $\#\text{-NP}$ can be logically characterized by independence logic and existential second-order logic, whereas dependence logic and inclusion logic give rise to subclasses of $\#\text{-NP}$ and $\#P$, respectively. Our main technical result shows that the problem of counting satisfying assignments for monotone Boolean Σ_1 -formulae is $\#\text{-NP}$ -complete as well as complete for the function class generated by dependence logic.

3.27 Finite-State Dependence

Dietmar Berwanger (CNRS, ENS Paris-Saclay, FR)

License  Creative Commons BY 3.0 Unported license
© Dietmar Berwanger

The use of information in games hinges on the notion of dependence. On the one hand, the actions of a player are constrained to depend on the information to which he has access. In turn, the consequence of an action can convey information to another player and thus loosen her respective information constraints.

Scenarios that involve successive (communication) actions tend to be difficult to analyse. Dependence logic appears as a suitable formalism to address such intricacies, at least in the specific case of coordination problems, which ask whether there exists a joint strategy that is successful.

If the information states are elements of a finite structure, and the objective function is first-order definable, the formulation of coordination problems in dependence logic is straightforward. However, in the more challenging settings of iterated games, information states correspond to sequences of observations, so we need to reason about infinite domains.

In the talk, we outline an interpretation of dependence logic in automatic structures with teams that admit finite-state representations. Essentially, the second-order objects involved in existential quantification and disjunction range over regular sets, and dependency relations are restricted to finite-state functions. We suggest a parametrisation of the semantics to ensure decidability via interpretation into the existential monadic theory of trees.

The perspective of this project is twofold: (1) to develop a suitable formalism for reasoning about information and coordination on the basis of dependence logic, and (2) to identify fragments of the logic that are decidable on automatic structures by retro-engineering positive results from automata theory and games with imperfect information.

3.28 The exact status of database semantics?

Joachim Biskup (TU Dortmund, DE)

License © Creative Commons BY 3.0 Unported license
© Joachim Biskup

Joint work of Joachim Biskup, Piero A. Bonatti

Main reference Joachim Biskup, Piero A. Bonatti: “Controlled query evaluation with open queries for a decidable relational submodel”, *Ann. Math. Artif. Intell.*, Vol. 50(1-2), pp. 39–77, 2007.

URL <https://doi.org/10.1007/s10472-007-9070-5>

Standard textbooks on relational database theory, as many original publications, usually assume that first-order logic can be employed for defining formal semantics for query answering, data dependencies and similar concepts. Regarding models, the authors of such work sometimes specify more precisely whether they have finite model theory or general model theory (allowing models of any cardinality) in mind. Accordingly, they consider either finite entailment or general entailment.

However, many works about relational database theory actually employ implicitly (or sometimes also explicitly) what we call database model theory

- on the syntactic layer, there is an infinite supply of constant symbols (0-ary function symbols);
- on the semantic layer, only Herbrand models with the fixed *infinite* universe consisting of the supplied constants on the one hand and only *finitely* many positively interpreted atomic sentences on the other hand, together with the pertinent unique names axioms for the constants, are considered.

The open problem then is the following: What is the exact status of the resulting database model semantics, in comparison to finite model semantics and general model semantics?

In many contexts the problem might be irrelevant, for example when studying safe and domain-independent queries. In some other contexts, however, distinguishing between the three types of semantics might be crucial, in particular when dealing with inference control where pure finite models could enable combinatorial reasoning. Moreover, database model semantics generate unusual tautologies. For example, for an open formula $f(x)$ – with x denoting the free variables – that is safe and domain-independent when seen as a query, the sentence $(\exists x)\neg f$ is true in all database models!

In our own work on inference control for open relational queries, see Section 2 of the main reference, we have exhibited a sufficient condition for the three kinds of semantics coinciding, introducing out-of-active-domain axioms besides the well-known unique names axioms. Though this condition has been helpful for a specific task, the exact relationship between the three semantics remains open.

Participants

- Leopoldo Bertossi
Carleton University – Ottawa,
CA & RelationalAI Inc., CA
- Dietmar Berwanger
CNRS, ENS Paris-Saclay, FR
- Meghyn Bienvenu
University of Bordeaux, FR
- Joachim Biskup
TU Dortmund, DE
- Katrin M. Dannert
RWTH Aachen, DE
- Anuj Dawar
University of Cambridge, GB
- Arnaud Durand
University Paris-Diderot, FR
- Fredrik Engström
University of Göteborg, SE
- Bernd Finkbeiner
Universität des Saarlandes, DE
- Floris Geerts
University of Antwerp, BE
- Erich Grädel
RWTH Aachen, DE
- Gianluca Grilletti
University of Amsterdam, NL
- Miika Hannula
University of Helsinki, FI
- Lauri Hella
Tampere University, FI
- Åsa Hirvonen
University of Helsinki, FI
- Matthias Hoelzel
RWTH Aachen, DE
- Benny Kimelfeld
Technion – Haifa, IL
- Phokion G. Kolaitis
University of California –
Santa Cruz, US
- Juha Kontinen
University of Helsinki, FI
- Paris Koutris
University of Wisconsin –
Madison, US
- Sebastian Link
University of Auckland, NZ
- Martin Lück
Leibniz Universität Hannover,
DE
- Yasir Mahmood
Leibniz Universität Hannover,
DE
- Arne Meier
Leibniz Universität Hannover,
DE
- Magdalena Ortiz
TU Wien, AT
- Martin Otto
TU Darmstadt, DE
- Eric J. Pacuit
University of Maryland –
College Park, US
- Henri Prade
University of Toulouse, FR
- David J. Pym
University College London, GB
- Raine Rönholm
Tampere University, FI
- Dan Suciu
University of Washington –
Seattle, US
- Val Tannen
University of Pennsylvania –
Philadelphia, US
- Bernhard Thalheim
CAU Kiel, DE
- Jouko Väänänen
University of Helsinki, FI
- Jan Van den Bussche
Hasselt University, BE
- Jonni Virtema
Hasselt University, BE
- Heribert Vollmer
Leibniz Universität Hannover,
DE
- Jef Wijsen
University of Mons, BE
- Richard Wilke
RWTH Aachen, DE
- Fan Yang
University of Helsinki, FI



Conditional Logics and Conditional Reasoning: New Joint Perspectives

Edited by

Guillaume Aucher¹, Paul Egré², Gabriele Kern-Isberner³, and Francesca Poggiolesi⁴

1 University of Rennes 1, CNRS, FR, guillaume.aucher@irisa.fr

2 CNRS, ENS, PSL University, FR, paul.egre@ens.fr

3 TU Dortmund, DE, gabriele.kern-isberner@cs.tu-dortmund.de

4 CNRS, University of Paris 1, ENS, FR, poggiolesi@gmail.com

Abstract

In the last decades, with the emergence of artificial intelligence, a large number of logics called conditional logics have been introduced to model our conditional reasoning captured by so-called conditionals, which are statements of the form ‘if A then B’. More recently, conditional reasoning has also come under scrutiny by psychologists, yet with more pragmatic and empirical considerations. The main objective of this seminar was to provide an opportunity for these different communities working on that topic to meet and reinforce their ties. We focused on three specific issues. First, we investigated how people’s intuitions about ‘counterpossibles’ can be understood empirically and classified with respect to the theoretical accounts of conditional logics. Second, we reconsidered the various semantics of system P and we wondered to which extent pragmatics plays a role in the relevance relation between the antecedent and the consequent of a conditional. Third, we strove to apply the recent advances in proof theory and correspondence theory to conditional logics. These working groups were preceded by short talks and tutorials.

Seminar January 13–16, 2019 – <https://www.dagstuhl.de/19032>

2012 ACM Subject Classification Artificial intelligence → philosophical/theoretical foundations of artificial intelligence, Artificial intelligence → knowledge representation and scheduling, Mathematical logic and formal languages → proof theory, Mathematical logic and formal languages → modal and temporal logic

Keywords and phrases Commonsense reasoning, conditionals, correspondence theory, proof theory, psychology of reasoning

Digital Object Identifier 10.4230/DagRep.1.9.47

Edited in cooperation with Mario Günther

1 Executive Summary

Guillaume Aucher (University of Rennes 1, CNRS, FR, guillaume.aucher@irisa.fr)

Paul Egré (CNRS, ENS, PSL, FR, paul.egre@ens.fr)

Gabriele Kern-Isberner (TU Dortmund, DE, gabriele.kern-isberner@cs.tu-dortmund.de)

Francesca Poggiolesi (CNRS, University of Paris 1, ENS, FR, poggiolesi@gmail.com)

License © Creative Commons BY 3.0 Unported license

© Guillaume Aucher, Paul Egré, Gabriele Kern-Isberner, Francesca Poggiolesi

Logic in the first half of the 20th century has been mostly concerned with mathematical reasoning and providing a unified framework for the foundations of mathematics. In the second half of the 20th century, with the emergence of artificial intelligence, new formalisms have been introduced to model kinds of inference closer to everyday life.



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

Conditional Logics and Conditional Reasoning: New Joint Perspectives,

Dagstuhl Reports, Vol. 9, Issue 1, pp. 47–66

Editors: Guillaume Aucher, Paul Egré, Gabriele Kern-Isberner and Francesca Poggiolesi



DAGSTUHL
REPORTS Dagstuhl Reports

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

“Commonsense reasoning”, the reasoning that humans perform in everyday life, is significantly different from the reasoning of mathematicians, which has been the object of study of (mathematical) logic for a long time. It is very rich and includes different kinds of reasoning, such as counterfactual reasoning, default reasoning or uncertain and plausible reasoning. Commonsense reasoning is often captured by means of conditionals, which are sentences of the form ‘if A then B ’. These conditionals can also be of various kinds: counterfactual, indicative, or subjunctive. The benefits of conditionals for formalizing commonsense reasoning are basically twofold: first, they can encode reasoning patterns of various types if one chooses suitable semantics or calculi, and second, they provide a common syntactic element that can be used to relate and compare the different kinds of commonsense reasoning as well as the mathematical reasoning.

Conditionals are also studied in the psychology of reasoning, which has recently witnessed a new wave of work. In particular, an effort to confront semantic frameworks with empirical results has been made. In parallel, a number of mathematical advances have been made in modal logic, an area closely related to conditional logics. However, the techniques developed in modal logic with respect to proof theory and correspondence theory have not fully been applied to the conditional logics introduced in artificial intelligence and philosophy. The main objective of this seminar was to provide an opportunity for computer scientists, logicians, psychologists, linguists and philosophers working on that topic to meet and reinforce their ties over several days in the Dagstuhl castle.

We focused on three specific issues which were discussed and worked out in three different working groups. First, we investigated how people’s intuitions about ‘counterpossibles’ can be understood empirically and classified thanks to the theoretical accounts of conditional logics. Second, we reconsidered the various semantics of the basic system P and wondered to which extent pragmatics plays a role in the relevance relation between the antecedent and the consequent of a conditional. Third, we strove to apply the recent advances in proof theory and correspondence theory to conditional logics. These three topics correspond respectively to the working groups “Investigating people’s intuitions about counterpossibles” (Section 4.1), “The semantics of conditionals” (Section 4.2) and “Correspondence theory and proof theory for conditional logics” (Section 4.3).

These working group discussions were preceded by 13 short talks and 3 tutorials: “Semantics of Conditionals” (by Graham Priest), “Proof Theory of Conditionals” (by Nicola Olivetti) and “The psychology of Indicative Conditionals” (by Karolina Krzyzanowska). These talks and tutorials are summarized in Section 3.

2 Table of Contents

Executive Summary

Guillaume Aucher, Paul Egré, Gabriele Kern-Isberner, Francesca Poggiolesi 47

Summary of Talks and Tutorials

| | |
|--|----|
| Representing and Reasoning with Conditionals: What Cognitive Neuroscience has (not (yet)) Taught Us <i>Giosué Baggio</i> | 51 |
| Conditionals: the Three-valued Approach <i>Didier Dubois</i> | 51 |
| Three-valued Conditionals and Three-valued Consequence <i>Paul Egré</i> | 52 |
| Proper Display Calculi for Conditional Logics via Multi-Type Correspondence Theory <i>Giuseppe Greco</i> | 53 |
| A Strengthened Ramsey Test Semantics for Missing-Link Conditionals <i>Mario Günther</i> | 54 |
| Connexive Logic and Conditional Logics <i>Andreas Kapsner</i> | 54 |
| Worlds are not Enough <i>Stephan Kaufmann</i> | 55 |
| A Common Semantic Base for Reasoning with Conditionals <i>Gabriele Kern-Isberner</i> | 55 |
| The Psychology of Indicative Conditionals (TUTORIAL) <i>Karolina Krzyzanowska</i> | 56 |
| Proof Theory of Conditional Logic (TUTORIAL) <i>Nicola Olivetti</i> | 57 |
| Conditionals: a Logician’s Perspective (TUTORIAL) <i>Graham Priest</i> | 57 |
| Difference-making Conditionals and the Relevant Ramsey Test <i>Hans Rott</i> | 58 |
| The Dialogical Entailment Task <i>Niels Skovgaard Olsen</i> | 58 |
| There Isn’t One Logic of Human Reasoning: What does ‘Pragmatic (Logical) Pluralism’ Mean for the Cognitive Science of Human Reasoning ? <i>Keith Stenning</i> | 59 |
| From Defeasible Conditionals to Preferential Modalities and Beyond <i>Ivan José Varzinczak</i> | 60 |
| On the Nature of Nonmonotonic Reasoning – Some Formal Clarifications <i>Emil Weydert</i> | 60 |

Working Groups

| | |
|--|----|
| Investigating People’s Intuitions about Counterpossibles <i>Nicole Cruz de Echeverria Loebell, Giosu  Baggio, Andreas Herzig, Andreas Kapsner, Karolina Krzyzanowska, Francois Olivier, Graham Priest, Keith Stenning, and Jakub Szymanik</i> | 61 |
| The Semantics of Conditionals <i>Paul Egr , Didier Dubois, Mario G nther, Stefan Kaufmann, Gabriele Kern-Isberner, Eric Raidl, Hans Rott, Niels Skovgaard Olsen, and Emil Weydert</i> | 61 |
| Correspondence Theory and Proof Theory for Conditional Logics <i>Guillaume Aucher, Christoph Beierle, Giovanni Casini, Marianna Girlando, Giuseppe Greco, Nicola Olivetti, and Ivan Jos  Varzinczak</i> | 64 |
| Participants | 66 |

3 Summary of Talks and Tutorials

3.1 Representing and Reasoning with Conditionals: What Cognitive Neuroscience has (not (yet)) Taught Us

Giosué Baggio (NTNU – Trondheim, NO)

License  Creative Commons BY 3.0 Unported license
© Giosué Baggio

In this talk I gave a flash review of EEG and fMRI experiments on conditional reasoning, calling attention to the theoretical issues addressed or raised by these studies: the mental models vs mental logic debate, dual-process accounts of reasoning, the separability of logical inference from other forms of inference, and the nature of the deductive process. I pointed out some methodological problems with some of these studies, and I argued that research on how conditionals are represented in the mind/brain (as syntactic and semantic objects) should serve as a prerequisite and foundation for research on conditional reasoning proper. I finally presented work from our group and associated labs engaging with this issue.

3.2 Conditionals: the Three-valued Approach

Didier Dubois (University of Toulouse, FR)

License  Creative Commons BY 3.0 Unported license
© Didier Dubois

Joint work of Didier Dubois, Henri Prade

Main reference Didier Dubois, Henri Prade: “Conditional objects as nonmonotonic consequence relationships” in IEEE Transactions on Systems, Man, and Cybernetics, vol. 24, no. 12, pp. 1724–1740, Dec. 1994.

URL <https://doi.org/10.1109/21.328930>

One question that has bothered philosophers such as Lewis and Stalnaker in the 1970’s is the difference between a conditional in Boolean logic (such as the material implication) and the conditional that appears in a conditional probability. In the 1930’s, De Finetti [1] had already suggested the answer. A conditional is a three-valued proposition, a more faithful representation of an if-then rule than a material implication. The idea is just to distinguish between possible worlds that are examples of a rule, those that are counterexamples, and finally situations where the rule does not apply [2]. Under this view a conditional is a pair of nested Boolean events whose probabilities fully characterize the conditional probability. It can be captured in a three-valued logic of conditional statements, that can be combined by Sobocinski’s conjunction (equivalent to Adams’ quasi-conjunction) [3]. The syntax and axioms of this logic are precisely the ones of System P of Kraus Lehman and Magidor [12], which is sound and complete with respect to this three-valued semantics. This is the simplest semantics for this non-monotonic logic; see [12, 6, 7, 8, 9] for other semantics. The works of G. Kern-Isberner [11] can be seen as akin to this tradition. Then the probability of such conditionals is indeed the conditional probability. There are two ways of defining such conditionals, one using a Boolean version of Bayes rule, and the other more explicit one, as a pair of Boolean events. The two definitions yield the same definition for conditional probability. But the two definitions differ for other set functions such as belief functions and possibility measures [4, 10]. These two forms of conditioning correspond to distinct tasks: one is for question answering based on evidence, the other is the revision of uncertain beliefs [5].

References

- 1 B. De Finetti. La logique de la probabilité. In Actes du Congrès International de Philosophie Scientifique (1935), Paris, France, 1936. Hermann et Cie Editions.
- 2 D. Dubois, I.R. Goodman, P.G. Calabrese: Special Issue on Conditional Event Algebra, IEEE Trans. on Systems, Man and Cybernetics, 24(12) 1994.
- 3 D. Dubois, H. Prade: Conditional objects as nonmonotonic consequence relationships. pp. 1724–1740 of this special issue
- 4 D. Dubois, H. Prade, P. Smets: Representing partial ignorance. IEEE Trans. on Systems, Man and Cybernetics, 26(3), 1996, 361–377.
- 5 D. Dubois, H. Prade: Focusing vs. Belief Revision: A Fundamental Distinction When Dealing with Generic Knowledge. Proc. ECSQARU-FAPR, Lecture Notes in Computer Science 1244, Springer 1997: 96–107
- 6 S. Benferhat, D. Dubois, H. Prade: Nonmonotonic reasoning, conditional objects and possibility theory. Artificial Intelligence Journal, 92, 259–276, 1997.
- 7 S. Benferhat, D. Dubois, H. Prade, Possibilistic and standard probabilistic semantics of conditional knowledge bases, Journal of Logic and Computation, 9, 873–895, 1999.
- 8 D. Dubois H. Fargier, H. Prade, Acceptance, Conditionals, and Belief Revision, In: G. Kern-Isberner et col. (Eds.) Conditionals, Information, and Inference, Lecture Notes in Artificial Intelligence, Vol. 3301, Springer, 2005, p. 38–58.
- 9 D. Dubois, H. Prade: Non-Monotonic Reasoning and Uncertainty Theories. In : Nonmonotonic Reasoning. Essays Celebrating its 30th Anniversary.. Gerhard Brewka, Victor Marek, Miroslaw Truszczyński (Eds.), College Publications, p. 141–176, Vol. 31, Studies in Logic, 2011.
- 10 D. Dubois, T. Denoeux: Conditioning in Dempster-Shafer Theory: Prediction vs. Revision. Belief Functions 2012: Advances in Intelligent and Soft Computing 164, Springer 2012, 385–392.
- 11 G. Kern-Isberner: Conditionals in Nonmonotonic Reasoning and Belief Revision – Considering Conditionals as Agents. Lecture Notes in Computer Science 2087, Springer 2001
- 12 S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence, 44:167–207, 1990.

3.3 Three-valued Conditionals and Three-valued Consequence

Paul Egré (CNRS, ENS, PSL University)

License  Creative Commons BY 3.0 Unported license
© Paul Egré

Logical consequence is standardly defined as the preservation of designated values from premises to conclusion in an argument. In recent years, some attention has been given to so-called **mixed** consequence relations, in which designated values are allowed to vary between premises and conclusion in an argument (viz. Cobreros, Egré, Ripley, van Rooij [6], Smith [11], Zardini [12]). In this presentation I report on two lines of ongoing work dealing with the typology of three-valued conditional operators, the first pursued with Emmanuel Chemla, the second with Lorenzo Rossi and Jan Sprenger. The first line of inquiry concerns the extraction of conditional operators that internalize three-valued (and more generally, many-valued) **intersective mixed** consequence relations (see Chemla, Egré, Spector [3], Chemla and Egré [4, 5]). The second concerns the selection of an adequate notion of validity for the family of three-valued de Finettian conditional operators (de Finetti [8], Reichenbach [10]), which

take the value of the consequent when the antecedent is true, and the value indeterminate when the antecedent is false. Specific attention is given to a variant of de Finetti's table introduced independently by Cooper [7] and Cantwell [2], and on whether it is better behaved if it is asked that a conditional operator **internalize** the consequence relation (in the sense of Avron [1]).

References

- 1 Avron, A. (1991). Natural 3-valued logics—Characterization and proof theory, *The Journal of Symbolic Logic*, 56(1): 276–294.
- 2 Cantwell, J. (2008). The logic of conditional negation. *Notre Dame Journal of Formal Logic*, 49(3):245–260.
- 3 Chemla, E., Egré, P., and Spector, B. (2017). Characterizing Logical Consequence in many-valued logic. *Journal of Logic and Computation*, 27(7):2193–2226.
- 4 Chemla, E. and Egré, P. (2018). From Many-Valued Consequence to Many-Valued Connectives. *Synthese* (forthcoming).
- 5 Chemla, E. and Egré, P. (2018). Suszko's Problem: Mixed Consequence and Compositionality. Manuscript available on arXiv.org: <https://arxiv.org/abs/1707.08017>.
- 6 Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2012). Tolerant, Classical, Strict. *Journal of Philosophical Logic* 41(2):347–385.
- 7 Cooper, W. S. (1968). The propositional logic of ordinary discourse. *Inquiry*, 11:295–320.
- 8 De Finetti, B. (1936). La logique de la probabilité. *Actes du congrès international de philosophie scientifique*, 4:1–9.
- 9 Egré, P., Rossi, L., and Sprenger, J. (2019). De Finettian Logics of Indicative Conditionals. Manuscript available on arXiv (under review).
- 10 Reichenbach, H. (1944). *Philosophic foundations of quantum mechanics*. University of California Press.
- 11 Smith, N. (2008). *Vagueness and Degrees of Truth*, Oxford University Press, Oxford.
- 12 Zardini, E. (2008). A model of tolerance. *Studia Logica*, 90(3):337–368.

3.4 Proper Display Calculi for Conditional Logics via Multi-Type Correspondence Theory

Giuseppe Grecco (University of Utrecht, NL)

License  Creative Commons BY 3.0 Unported license
© Giuseppe Grecco

Connections between correspondence phenomena and proof theory have been seminaly observed and exploited by Marcus Kracht, in the context of his characterization of modal axioms which can be effectively transformed into ‘analytic’ structural rules of display calculi. Applying insights from unified correspondence theory, Kracht's results were extended to the setting of DLE-logics (logics the algebraic semantics of which is based on bounded distributive lattices) characterizing the space of ‘properly displayable DLE-logics’. In a series of co-authored papers, I contributed to extend the boundaries of this line of research in structural proof theory to a number of logics captured by axioms that are not analytic in the original language. In my presentation, I analyzed the features of well known (non-analytic) axioms for conditional logics challenging a unified logical framework, and I suggested possible directions of research.

3.5 A Strengthened Ramsey Test Semantics for Missing-Link Conditionals

Mario Günther (*Universität Regensburg, DE*)

License © Creative Commons BY 3.0 Unported license
© Mario Günther

Joint work of Mario Günther, Holger Andreas

Main reference Holger Andreas, Mario Günther: “On the Ramsey Test Analysis of ‘Because’”, in *Erkenntnis*, June 2018.

URL <https://doi.org/10.1007/s10670-018-0006-8>

To accept a conditional ‘If A then B ’, so suggest Douven [2] and Krzyżanowska et al. [3], requires a relation of relevance between its antecedent A and its consequent B . Conditionals that miss a relevant link, such as “If Lund is a town in Sweden, Munich is a town in Germany” sound odd because there is no apparent connection between the antecedent and the consequent. Douven [2, p.1542] observes that “none of the currently popular semantics” elevates a relevance relation to be a necessary condition for a conditional to hold. In this talk, we aim to remedy this situation. Inspired by Rott [4], we strengthen the Ramsey Test by a suspension of judgment: accept ‘If A then B ’ iff, after suspending judgment on A and B , you can infer B by assuming A . The suspension of judgment creates a context of the remaining beliefs. If A , together with the remaining beliefs, is sufficient to infer B , we accept the conditional; otherwise we do not. Andreas and Günther [1] have shown that this strengthened Ramsey Test gives rise to a new semantics for conditionals that requires A to be inferentially relevant for B . We argue that the strengthened Ramsey Test semantics can solve the challenge posed by missing-link conditionals.

References

- 1 Andreas, H. and Günther, M. (2018). On the Ramsey Test Analysis of ‘Because’. *Erkenntnis*, <https://doi.org/10.1007/s10670-018-0006-8>.
- 2 Douven, I. (2017). How to account for the oddness of missing-link conditionals. *Synthese* 194(5): 1541–1554.
- 3 Krzyżanowska, K., Collins, P. J., and Hahn, U. (2017). Between a conditional’s antecedent and its consequent: Discourse coherence vs. probabilistic relevance. *Cognition* 164: 199–205.
- 4 Rott, H. (1986). Ifs, Though, and Because. *Erkenntnis* 25(3): 345–370.

3.6 Connexive Logic and Conditional Logics

Andreas Kapsner (*LMU München, DE*)

License © Creative Commons BY 3.0 Unported license
© Andreas Kapsner

Connexive logics are non-classical logics that validate the following intuitively appealing principles:

Aristotle: $\neg(A \rightarrow \neg A)$ and $\neg(\neg A \rightarrow A)$ are valid.

Boethius: $(A \rightarrow B) \rightarrow \neg(A \rightarrow \neg B)$ and $(A \rightarrow \neg B) \rightarrow \neg(A \rightarrow B)$ are valid.

These are not valid in classical logic, and in fact, they aren’t valid in most well-known non-classical logics, either. (A good place to get a first orientation about the topic is the SEP entry “Connexive Logic”.)

In this short talk, I will comment on the relationship between connexive logics and conditional logics in the Lewis/Stalnaker family of theories. In particular, I will be interested in the philosophical underpinnings of these two large projects and in how much these underpinnings intersect. Though it has always been clear that there seems to be some connection here, it has, I believe, not yet been established what that connection is, precisely. I will propose a view of connexivity (drawing on earlier work) that not only fits well to the philosophical discussion about conditional logics, but is also able to shed new light on topics in that discussion, such as the dispute about the Law of Conditional Non-Contradiction.

3.7 Worlds are not Enough

Stephan Kaufmann (University of Connecticut, USA)

License  Creative Commons BY 3.0 Unported license
© Stephan Kaufmann

The thesis that the probability of a conditional is the corresponding conditional probability – henceforth *The Thesis* – has long intrigued philosophers and enjoys wide and growing support in psychology. It seems to hold in a wide range of situations, though not without exception. However, despite its merits, no semantic theory based on the *The Thesis* has established itself in the mainstream. There are two reasons for this. Firstly, Lewis’s triviality results and much subsequent work showed that any such semantic theory would have to be at odds with certain entrenched assumptions about the proper formal framework, and logicians starting with Lewis himself have been queasy about the radical changes that would be required to resolve this tension. Secondly, and relatedly, it is unclear how a theory incorporating *The Thesis*, if one can be given, would be integrated with semantic approaches to other linguistic forms, or for that matter with the wider context of epistemological and metaphysical theories with which familiar possible-worlds models interface so seamlessly.

But it has been known for some time that a semantic theory building on *The Thesis* can be developed, and in view of the renewed interest in probabilistic semantics in linguistics, philosophy, psychology and artificial intelligence, this is a good time to work it out and explore the consequences. My work in this area builds on the “Bernoulli Models” pioneered by van Fraassen and further developed by Stalnaker and Jeffrey. Specifically, I am interested in (i) the predictions of this framework about compounds with embedded conditionals (and theoretical knobs to turn where those predictions clash with intuition); (ii) extensions to linguistic forms other than ‘if-then’ sentences, and to various forms of context dependence; and (iii) models of belief and belief dynamics.

3.8 A Common Semantic Base for Reasoning with Conditionals

Gabriele Kern-Isberner (TU Dortmund, DE)

License  Creative Commons BY 3.0 Unported license
© Gabriele Kern-Isberner

There is a large variety of different conditional logics with different semantics, and beyond that, there are general semantic frameworks to evaluate conditional statements, like probabilities, possibilities, or plausibilities. In this short talk, I address the problem of what general principles guide the derivation of “new” conditionals from a conditional knowledge base. I

present the basics of the theory of conditional structures which is built upon De Finetti's 3-valued conditionals, and which becomes effective for inductive reasoning and belief revision via the principle of conditional preservation. This principle can be linked to various semantic frameworks, in particular, probabilities, possibilities, plausibilities, and purely qualitative preorders.

Therefore, conditional structures provide a versatile algebraic framework for reasoning with conditionals for various tasks:

- probabilistic reasoning via the principle of maximum entropy;
- iterated belief revision covering the seminal postulates by [Darwiche & Pearl, 1997];
- nonmonotonic reasoning from conditional knowledge bases as an alternative to Pearl's system Z (with better results in many cases);
- extracting background knowledge from the statistical outcomes of empirical studies that helps explaining people's reasoning behaviour.

3.9 The Psychology of Indicative Conditionals (Tutorial)

Karolina Krzyzanowska (*University of Amsterdam, NL*)

License  Creative Commons BY 3.0 Unported license
© Karolina Krzyzanowska

Indicative conditionals play a central role in reasoning. Unsurprisingly then, a lot of psychology of reasoning research has been devoted to the question of how people interpret conditional sentences and what kind of inferences they draw from them. Starting with the ancestor of all reasoning tasks, Wason's card selection problem (1966), I present some of the most important findings about people's interpretation of conditionals. Finally, I discuss three recent experiments that highlight the significance of a relevance relation between the conditionals' antecedents and consequents.

References

- 1 Barrouillet, P., Grosset, N., & Lecas, J.F. (2000). "Conditional reasoning by mental models: Chronometric and developmental evidence". *Cognition* 75(3): 237–266.
- 2 Byrne, R. M. J. (1989). "Suppressing valid inferences with conditionals". *Cognition* 31: 61–83.
- 3 Douven, I. and Verbrugge, S. (2010). "The Adams family". *Cognition*, 117: 302–318.
- 4 Douven, I., Elqayam, S, Singmann, H., van Wijnbergen-Huitink, J. (2018). "Conditionals and inferential connections: A hypothetical inferential theory". *Cognitive Psychology* 101: 50–81.
- 5 Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). "Conditionals and conditional probability". *Journal of Experimental Psychology: Learning, Memory and Cognition* 29: 321–355.
- 6 Oaksford, M., & Chater, N. (1994). "A rational analysis of the selection task as optimal data selection". *Psychological Review* 101: 608–631.
- 7 Oaksford, M., & Chater, N. (2003). "Conditional probability and the cognitive science of conditional reasoning". *Mind and Language* 18(4): 359–379.
- 8 Skovgaard-Olsen, N., Singmann, H., and Klauer, K. C. (2016). "The relevance effect and conditionals". *Cognition*, 150: 26–36
- 9 Wason, P.C. (1966). "Reasoning". In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, UK: Penguin.

3.10 Proof Theory of Conditional Logic (Tutorial)

Nicola Olivetti (Aix-Marseille University, FR)

License  Creative Commons BY 3.0 Unported license
© Nicola Olivetti

Many systems of conditional logics have been proposed the last 50 years. They can be naturally grouped in three families determined by their semantics, namely Basic Conditional Logics, Preferential Conditional Logics, and Lewis' Logics of Counterfactuals. If semantics and axiomatization of each system are both well understood, their proof theory, in the form of sequent calculi, is not as developed as the one of other families of logics. In this tutorial I first propose the main requirements or properties of proof systems, and then I present analytic sequent calculi for each family of conditional logics at the state of the art.

References

- 1 R. Alenda, N. Olivetti, and G. L. Pozzato, Nested Sequents Calculi for Normal Conditional Logics. *Journal of Logic and Computation*, 26 (1): 7–50, 2016.
- 2 H. C. M. de Swart, A Gentzen- or Beth-type system, a practical decision procedure and a constructive completeness proof for the counterfactual logics VC and VCS. *Journal of Symbolic Logic*, 48, 1–20, 1983.
- 3 I. P. Gent, A sequent or tableaux-style system for lewis's counterfactual logic VC. *Notre Dame Journal of Formal Logic*, 33(3): 369–382, 1992.
- 4 M. Giraldo, B. Lellmann, N. Olivetti, G. L. Pozzato, Standard Sequent Calculi for Lewis' Logics of Counterfactuals. JELIA 2016, Logics in Artificial Intelligence, LNAI 10021, 2016: pp. 272–287, 2016.
- 5 M. Giraldo, B. Lellmann, N. Olivetti, G. L. Pozzato, Hypersequent Calculi for Lewis' Conditional Logics with Uniformity and Reflexivity. TABLEAUX 2017, Automated Reasoning with Analytic Tableaux and Related Methods, Vol. LNCS 10501, pp. 131–148, 2017.
- 6 B. Lellmann, D. Pattinson, Sequent Systems for Lewis' Conditional Logics. *JELIA 2012*, 320–332, 2012.
- 7 S. Negri and N. Olivetti, A sequent calculus for preferential conditional logic based on neighbourhood semantics. *TABLEAUX 2015*, Vol LNCS 9323, pp. 115–134, 2015.
- 8 S. Negri and G. Sbardolini, Proof analysis for Lewis counterfactuals. *The Review of Symbolic Logic*, vol 9: 44–75, 2016.
- 9 D. Pattinson and L. Schröder, Generic modal cut elimination applied to conditional logics. *Logical Methods in Computer Science*, 7, 2011.
- 10 F. Poggiolesi, Natural Deduction Calculi and Sequent Calculi for Counterfactual Logics. *Studia Logica* 104(5): 1003–1036, 2016.

3.11 Conditionals: a Logician's Perspective (Tutorial)

Graham Priest (CUNY Graduate Center, USA)

License  Creative Commons BY 3.0 Unported license
© Graham Priest

In this talk, I will suggest that a conditional, $A > B$, is true iff B is true in all those worlds where A holds and where certain contextually determined information, imported from the actual world, also holds. I will explain a standard formal semantics and tableau proof system for such a view. I will then discuss how this view bears on three further issues. (1) Inferences that are formally invalid, but are contextually correct. (2) The distinction between indicative and subjunctive conditionals. (3) Conditionals with impossible antecedents.

References

- 1 Berto, F., and Jago, M. (2018), ‘Impossible Worlds, in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/impossible-worlds/>.
- 2 Priest, G., (2008), *Introduction to Non-Classical Logic*, Cambridge: Cambridge University Press, chs. 5 and 19.
- 3 Priest, G. (2018), ‘Some New Thoughts on Conditionals’, *Topoi* 27: 369–77.

3.12 Difference-making Conditionals and the Relevant Ramsey Test

Hans Rott (Universität Regensburg, DE)

License  Creative Commons BY 3.0 Unported license
© Hans Rott

This paper explores conditionals expressing that the antecedent makes a difference for the consequent. It employs a ‘relevantised’ version of the Ramsey Test for conditionals in the context of the classical theory of belief revision due to Alchourrón, Gärdenfors and Makinson (1985). The idea of this test is that the antecedent is relevant to the consequent in the following sense: a conditional is accepted just in case the consequent is accepted if the belief state is revised by the antecedent *and* fails to be accepted if the belief state is revised by the antecedent’s negation. The connective thus defined violates almost all of the traditional principles of conditional logic, but it obeys an interesting logic of its own.

The paper also gives the logic of an alternative version, the ‘Dependent Ramsey Test’ according to which a conditional is accepted just in case the consequent is accepted if the belief state is revised by the antecedent *and* is rejected (e.g., its negation is accepted) if the belief state is revised by the antecedent’s negation.

References

- 1 Igor Douven. *The epistemology of indicative conditionals: Formal and empirical approaches*. Cambridge University Press, Cambridge, 2016.
- 2 Luis Fariñas del Cerro and Andreas Herzig. Belief change and dependence. In Yoav Shoham (ed.), *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*. Morgan Kaufmann, San Francisco, CA, 1996, pp. 147–161.
- 3 Peter Gärdenfors. Conditionals and changes of belief. In I. Niiniluoto & R. Tuomela (eds.), *The Logic and Epistemology of Scientific Change*. Amsterdam: North Holland, 1978, pp. 381–404.
- 4 Hans Rott. Ifs, though and because. *Erkenntnis* 25 (1986), pp. 345–370.

3.13 The Dialogical Entailment Task

Niels Skovgaard Olsen (Universität Göttingen, DE)

License  Creative Commons BY 3.0 Unported license
© Niels Skovgaard Olsen

The goal of this talk is to present a novel experimental paradigm for probing the participants’ acceptance of entailments. In the psychology of reasoning, there has been a recent change from the use of deductive task instructions to probabilistic tasks instructions by Singmann & Klauer [1]. One side-effect of this change is that entailment judgments are no longer a primary focus of investigation although they arguably constitute the main source of data for semantic

theories [4]. This even holds for recent studies on p-validity within the new paradigm, which arguably probe probabilistic coherence in probability assignments to the premises and the conclusion rather than entailment judgments per se. Through the introduction of the Dialogical Entailment Task, we seek to remedy this predicament [2]. By adopting this task, the participants' acceptance of inferences with conditionals and negation operators are investigated and evidence is found that the participants do not accept the equivalence of wide and narrow scope negations of indicative conditionals, across relevance levels. As such, these results are in line with recent results on the interaction of conditionals with negation operators in probability judgments under manipulations of relevance [3].

References

- 1 Singmann, H., & Klauer, K. C. (2011). Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & Reasoning*, 17(3), 247–281.
- 2 Skovgaard-Olsen, N. (in review). The Dialogical Entailment Task.
- 3 Skovgaard-Olsen, N., Collins, P., Krzyzanowska, K., Hahn, U., and Klauer, C. K. (2019). Cancellation, Negation, and Rejection. *Cognitive Psychology*, 108, 42–71.
- 4 Winter, Y. (2016). *Elements of Formal Semantics*. Edinburgh: Edinburgh University Press.

3.14 There Isn't One Logic of Human Reasoning: What does 'Pragmatic (Logical) Pluralism' Mean for the Cognitive Science of Human Reasoning ?

Keith Stenning (University of Edinburgh, UK)

License © Creative Commons BY 3.0 Unported license
© Keith Stenning

Michiel van Lambalgen and I have been developing a multiple-logics framework for cognition for some years (Stenning & van Lambalgen 2008). The empirical part of this work has mostly focussed on modelling human nonmonotonic narrative reasoning in Logic Programming. This focus is required by the neglect (sometimes even denial) of this topic as reasoning. Recently however, we have turned to revisiting human monotonic reasoning (the classical logical syllogism) to show how much empirically richer this becomes if one takes the logic seriously. If instead of assuming that the conventional 'draw-a-conclusion-from-these-premisses' task invokes a classical logical goal in 'logically naive subjects', one compares it with a more obvious situation of dispute. This talk would attempt to motivate the study of betting against Harry-the-Snake on the validity of syllogisms, as a task more suitable for invoking classical logic in these subjects. And to give a flavour of some preliminary results.

3.15 From Defeasible Conditionals to Preferential Modalities and Beyond

Ivan José Varzinczak (Artois University, FR)

License  Creative Commons BY 3.0 Unported license
© Ivan José Varzinczak

We investigate an aspect of defeasibility that has somewhat been overlooked by the non-monotonic reasoning community, namely that of defeasible modes of reasoning. These aim to formalise defeasibility of the traditional notion of necessity in modal logic, in particular of its different readings as action, knowledge and others in specific contexts, rather than defeasibility of conditional forms. Building on an extension of the preferential approach to modal logics, we introduce new modal operators with which to formalise the notion of defeasible necessity and distinct possibility, and that can be used to represent expected effects, refutable knowledge, and so on. We show how KLM-style conditionals can smoothly be integrated within our richer language. Moreover, we show that the computational complexity of the resulting framework remains in the same class as that of the underlying classical modal logic we start off with. Finally, we also show how our semantic constructions are fruitful in similarly structured logics such as description logics.

3.16 On the Nature of Nonmonotonic Reasoning – Some Formal Clarifications

Emil Weydert (University of Luxembourg, LU)

License  Creative Commons BY 3.0 Unported license
© Emil Weydert

We provide a general perspective on semantic-based approaches to default reasoning, pointing to several important – but hardly known – facts, as well as to desirable principles and standards. This includes the

- Interpretation of defaults as object-level expressions evaluated by suitable semantic structures fixing their acceptance
- Distinction – crucial and reflected by diverging logical properties – between object-level defaults ($a > b$) in the default base D , and meta-level inferential relationships $a \mid D b$ based on D
- Identification of 3 central postulates at the level of defeasible reasoning with defaults: (1) Irrelevance, (2) Boolean invariance, (3) Left Logical Equivalence for defaults
- Observation that ranking measures (rkm) are the simplest linear plausibility valuations correctly handling independence
- Fact that rational rkm-values are both necessary and sufficient for advanced default reasoning
- Insight that there is a necessary tradeoff between, on one hand, strong and intuitively adequate default inference relations (e.g. ME or JZ), and on the other hand, well-behaved conditional logics for the defaults themselves. In particular, it turns out that AND and RW (Right Weakening) are not compatible with the former.

4 Working Groups

4.1 Investigating People's Intuitions about Counterpossibles

Nicole Cruz de Echeverria Loebell (Birkbeck, University of London, UK), Giosué Baggio (NTNU – Trondheim, NO), Andreas Herzig (CNRS, FR), Andreas Kapsner (LMU München, DE), Karolina Krzyzanowska (University of Amsterdam, NL), Francois Olivier (ENS – Paris, FR), Graham Priest (CUNY Graduate Center, USA), Keith Stenning (University of Edinburgh, UK), Jakub Szymanik (University of Amsterdam, NL)

License © Creative Commons BY 3.0 Unported license
 © Nicole Cruz de Echeverria Loebell, Giosué Baggio, Andreas Herzig, Andreas Kapsner, Karolina Krzyzanowska, Francois Olivier, Graham Priest, Keith Stenning, and Jakub Szymanik

Our group started out discussing broader questions around counterpossibles, counterfactuals, contextualisation/framing, and related topics. But towards the end we narrowed our focus to counterpossibles, i.e. conditionals with an impossible antecedent, and to what intuitions people might have about these conditionals. We are planning an online experiment to assess this question. In the experiment, we seek to compare three accounts of the semantics of counterpossibles, which we refer to as vacuism, non-vacuism, and suspension-of-judgment. According to non-vacuism, people will judge some counterpossibles as true and others as false. For example, they will judge ‘If $1 + 2 = 4$, then $1 + 3 = 5$ ’ more often as true than ‘If $1 + 2 = 4$, then $1 + 3 = 147$ ’. According to vacuism, people will tend to judge any counterpossible as (vacuously) true. According to the suspension-of-judgment account, people will tend to suspend judgment when encountering a counterpossible, and so will tend to judge that the truth of any counterpossible cannot be determined. We will compare people’s judgments about counterpossibles with their judgments about corresponding possible conditionals, *e.g.* ‘If $1 + 2 = 3$, then $1 + 3 = 5$ ’. We will also explore potential differences in people’s judgments when the above conditionals are formulated in the subjunctive as opposed to indicative mood. Through this experiment we hope to gain information about how people understand and reason with counterpossibles, and to what extent the three theoretical accounts compared capture people’s intuitions about them.

4.2 The Semantics of Conditionals

Paul Egré (coordinator, ENS – Paris, FR), Didier Dubois (CNRS, FR), Mario Günther (Universität Regensburg, DE), Stefan Kaufmann (University of Connecticut – Storrs, US), Gabriele Kern-Isberner (TU Dortmund, DE), Eric Raidl (Universität Konstanz, DE), Hans Rott (Universität Regensburg, DE), Niels Skovgaard Olsen (Universität Göttingen, DE), Emil Weydert (University of Luxembourg, LU)

License © Creative Commons BY 3.0 Unported license
 © Paul Egré, Didier Dubois, Mario Günther, Stefan Kaufmann, Gabriele Kern-Isberner, Eric Raidl, Hans Rott, Niels Skovgaard Olsen, and Emil Weydert

Our group considered three main topics for discussion:

- (A) What is the relation between probabilistic semantics and possible-worlds semantics for conditionals?
- (B) How do truth-functional approaches to conditionals and intensional approaches compare?
- (C) Which role does relevance play as regards conditionals?

The discussion on (A) and (B) centered on the System P of non-monotonic logic, as laid out by Kraus et al. [11] (KLM) who claim it to be the “conservative core of a nonmonotonic reasoning system”. Under the restriction to non-nested conditionals, there are various ways to semantically describe the rules of System P: the probabilistic semantics of Adams [1]; a certain possible worlds semantics à la Stalnaker [17] and Lewis [12] or Chellas [6], namely Burgess semantics [5]; the three-valued approach of Dubois and Prade [9]; and, of course, the preferential models of KLM. It is no wonder, then, that System P is often taken to be a fundamental system of conditional logic. The discussion left open whether we can find translations between the various semantics without invoking System P. Another unsettled issue that came up was whether Dubois and Prade’s system really has the expressiveness of KLM’s system, or is in fact less expressive.

As regards (C), the question emerged whether a condition of relevance should constrain the semantics of conditionals, or whether considerations of relevance rather belong to their pragmatics only. It proved hard, to say the least, to find a clear demarcation between the semantics and pragmatics of conditionals. Different operationalizations of the notion of relevance were discussed, with specific attention to the probabilistic constraint that a conditional of the form $A > B$ is relevantly asserted provided $Pr(B | A) - Pr(B) > 0$ (as discussed by Douven [8, Chs. 4&5], with a recent variant considered by Crupi and Iacona [7]). Various qualitative alternatives were considered.

A paper presented by Hans Rott [15] during the Dagstuhl meeting, titled “Difference-making conditionals and the Relevant Ramsey Test” was given particular attention. The idea of this test is that the antecedent is relevant to the consequent in the following sense: a conditional is accepted just in case the consequent is accepted if the belief state is revised by the antecedent and fails to be accepted if the belief state is revised by the antecedent’s negation. Even though Rott does not conceive of the Relevant Ramsey Test as a compound of two object-language sentences, Eric Raidl suggested to interpret the proposal as follows: a conditional $A \gg B$ is a difference-making conditional provided the conjunction $(A > B) \wedge \neg(\neg A > B)$ is true or accepted. In fact, Raidl [13, 14] has already worked out that the relation of sufficient reason due to Spohn [18] can be spelled out by plugging the conditional based on the ranking-theoretic Ramsey Test into the conjunctive schema.

In the discussion, Mario Günther put forth the conjecture that the second conjunct expresses the pragmatic component of a difference-making conditional. A reason supporting the conjecture is that uttering ‘If A then B ’ explicitly states the first conjunct, while it leaves the second implicit. Niels Skovgaard-Olsen provided a reason against the conjecture based on the empirical work [16] investigating relevance effects found for the probability assessment of indicative conditionals. He argued that the empirical evidence is most consistent with making a positive delta- p value ($P(B | A) - P(B | \neg A) > 0$) a conventional implicature. A conventional implicature is not cancellable, does not arise on the basis of Gricean maxims of communication, and does not normally influence direct truth evaluations of conditionals. It is rather a second layer of meaning that is lexically encoded into conditionals, and as part of the conventional meaning of conditionals, it is a part of the semantic content of indicative conditionals. However, it remained contestable whether indicative conditionals carry a (probabilistic) counterfactual meaning arising from the second conjunct of difference-making conditionals.

In his Dagstuhl talk, Mario Günther delivered yet another strengthening of the Ramsey Test to spell out a relevance relation between the antecedent and the consequent. The idea is to accept ‘If A then B ’ iff, after suspending judgment on A and B , you can infer B by assuming A (see Andreas and Günther [2]). Andreas and Günther [3] provide an analysis of causation by plugging in this strengthened Ramsey Test conditional in the above conjunctive schema.

Among the tasks entertained as possible continuations of the discussion group, two main projects were put forth:

1. Elaborate a survey paper on bridges between semantic approaches giving system P, while making sure to not duplicate extant surveys, which provide translations between some semantics for P and semantics for PCL in the full language, e.g. Arló-Costa and Shapiro [4] and Friedman and Halpern [10].
2. Prepare a paper on the division of labor between semantics and pragmatics and the relation between inferential and difference-making conditionals.

References

- 1 Adams, E. (1966). Probability and the Logic of Conditionals, in *Aspects of Inductive Logic*, J. Hintikka and P. Suppes (eds.), pp. 265–316.
- 2 Andreas, H. and Günther, M. (2018). On the Ramsey Test Analysis of ‘Because’. *Erkenntnis*, <https://doi.org/10.1007/s10670-018-0006-8>.
- 3 Andreas, H. and Günther, M. (2019). Causation in Terms of Production. *Philosophical Studies*, <https://doi.org/10.1007/s11098-019-01275-3>.
- 4 Arló-Costa, H. and S. Shapiro (1992). Maps Between Conditional Logic and Non-monotonic Logic. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, B. Nebel, C. Rich, W. Swartout, (eds.), San Mateo, CA.: Morgan Kaufmann, pp. 553–565.
- 5 Burgess, J. (1981). Quick Completeness Proofs for Some Logics of Conditionals, *Notre Dame Journal of Formal Logic*, 22:76–84.
- 6 Chellas, B.F. (1975). Basic Conditional Logic, *Journal of Philosophical Logic*, 4:133–153.
- 7 Crupi, V. and Iacona, A. (2018, December). *Three ways of being non-material*. Manuscript.
- 8 Douven, I. (2016). *The epistemology of indicative conditionals: Formal and empirical approaches*. Cambridge University Press, Cambridge.
- 9 D. Dubois, H. Prade (1994). Conditional objects as nonmonotonic consequence relationships. *Special Issue on Conditional Event Algebra, IEEE Trans. on Systems, Man and Cybernetics*, 24(12):1724–1740.
- 10 Friedman, N. and Halpern, J. (2001). Plausibility measures and default reasoning. *Journal of the ACM*, 48:648–685.
- 11 Kraus, S., Lehmann, D. and Magidor. M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207.
- 12 Lewis, D. K. (1973). *Counterfactuals*. Blackwell Publishers.
- 13 Raidl, E. (2018a). Ranking semantics for doxastic necessities and conditionals, in *Logica Yearbook 2017*, Arazim, P. and Lávička, T. (eds), London: College Publications, pp. 223–238.
- 14 Raidl, E. (2018b). Completeness for counter-doxa conditionals – using ranking semantics, *The Review of Symbolic Logic*, <https://doi.org/10.1017/S1755020318000199>.
- 15 Rott, H. (2019, January). *Difference-making conditionals and the Relevant Ramsey Test*, Manuscript.
- 16 Skovgaard-Olsen, N., Collins, P., Krzyżanowska, K., Hahn, U. Klauer, K. C. (2019). Cancellation, negation, and rejection, *Cognitive Psychology*, 108:42–71, <https://doi.org/10.1016/j.cogpsych.2018.11.002>.
- 17 Stalnaker, R. (1968). A Theory of Conditionals. In Rescher, N. (ed.), *Studies in Logical Theory (American Philosophical Quarterly Monograph Series)*, pp. 98–112.
- 18 Spohn, W. (2015). Conditionals: A unifying ranking-theoretic perspective. *Philosophers’ Imprint*, 15(1), 1–30.

4.3 Correspondence Theory and Proof Theory for Conditional Logics

Guillaume Aucher (coordinator, University of Rennes 1, CNRS, FR), Christoph Beierle (Fern Universität in Hagen, DE), Giovanni Casini (University of Luxembourg, LU), Marianna Girlando (Aix-Marseille University, FR), Giuseppe Greco (Utrecht University, NL), Nicola Olivetti (Aix-Marseille University, FR), Ivan José Varzinczak (Artois University, FR)

License © Creative Commons BY 3.0 Unported license

© Guillaume Aucher, Christoph Beierle, Giovanni Casini, Marianna Girlando, Giuseppe Greco, Nicola Olivetti, and Ivan José Varzinczak

Our group dealt with the problem of finding out an appropriate semantics for the family of conditional logics in order to develop a uniform correspondence theory and proof systems for these logics.

Many different semantics have been defined in the literature for conditional logics, such as the selection function semantics, the preferential semantics or the sphere semantics. Selection function models are sound and complete with respect to \mathbb{CK} , the basic system of conditional logics, and can be extended to the whole family by specifying properties of the selection function. Thus, strictly speaking, a unified semantics for the family of conditional logic already exists. However, as underlined in [4], this semantics is not enough to define a proof system for the strongest systems, as the properties added are not informative enough to define rules. Moreover, the selection function semantics is not really suitable for developing a correspondence theory between semantics and proof systems. Hence, we looked at some richer models capturing \mathbb{CK} and extensions.

Our first attempt was to consider the Routley-Meyer semantics as a unifying semantics for conditionals, because a correspondence theory based on this semantics can be developed from the correspondence theory for modal logic based on Kripke models [1]. This semantics, defined for relevance logics, adds to standard Kripke models a ternary relation R among worlds: $Rxyz$. The truth condition of the conditional operator in this class of models is the following:

$$x \Vdash A > B \text{ iff for all } z, y \in W, \text{ if } Rxyz \text{ and } y \Vdash A, \text{ then } z \Vdash B.$$

It is possible to construct a class-selection function model from a Routley-Meyer model by defining the class-selection function as follows:

$$y \in f(S, x) \text{ iff there exists } z \in S \text{ such that } Rxyz.$$

Similarly, the ternary relation of the Routley-Meyer semantics can be defined in terms of the class-selection as follows:

$$Rxyz \text{ iff there exists } S \subseteq W \text{ such that } z \in S, \text{ and } z \in f(S, x).$$

However, the conditional defined on the basis of this ternary semantics will yield in any case a *monotonic* inference relation. Moreover, this semantics validates as well the axiom (OR) of the conditional logic \mathbb{PCL} , which should not be valid in the weaker system \mathbb{CK} . So, the Routley-Meyer semantics cannot be used to represent the conditional operators in their full generality.

This led us to focus on a different class of models, namely preferential models, first defined for the conditional logic \mathbb{PCL} [2]. These models add to Kripke structures a preferential ordering among worlds, which can also be expressed in terms of a ternary relation. Our conjecture is that by dropping some properties of the relation we might capture logics weaker than \mathbb{PCL} and, in particular, \mathbb{CK} .

References

- 1 Guillaume Aucher. A new road towards universal logic? Research report, Université de Rennes 1, April 2017.
- 2 John P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84, 1981.
- 3 Nir Friedman and Joseph Y. Halpern. On the complexity of conditional logics. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR'94)*, pages 202–213. Morgan Kaufmann Pub, 1994.
- 4 Laura Giordano, Valentina Gliozzi, Nicola Olivetti, and Camilla Schwind. Tableau calculus for preference-based conditional logics: PCL and its extensions. *ACM Transactions on Computational Logic (TOCL)*, 10(3):21, 2009.
- 5 Marcus Kracht. Power and weakness of the modal display calculus. In *Proof theory of modal logic*, pages 93–121. Springer, 1996.
- 6 Donald Nute. *Topics in conditional logic*, volume 20. Springer Science & Business Media, Switzerland, 1980.

Participants

- Guillaume Aucher
University of Rennes 1,
CNRS, FR
- Giosué Baggio
NTNU – Trondheim, NO
- Christoph Beierle
Fern Universität in Hagen, DE
- Didier Dubois
CNRS, Toulouse, FR
- Giovanni Casini
University of Luxembourg, LU
- Nicole Cruz de
Echeverria Loebell
Birkbeck, University of
London, UK
- Paul Egré
ENS – Paris, FR
- Marianna Girlando
Aix-Marseille University, FR
- Giuseppe Greco
Utrecht University, NL
- Mario Günther
Universität Regensburg, DE
- Andreas Herzig
CNRS, Toulouse, FR
- Andreas Kapsner
LMU München, DE
- Stefan Kaufmann
University of Connecticut –
Storrs, US
- Gabriele Kern-Isberner
TU Dortmund, DE
- Karolina Krzyzanowska
University of Amsterdam, NL
- Nicola Olivetti
Aix-Marseille University, FR
- Francois Olivier
ENS – Paris, FR
- Graham Priest
CUNY Graduate Center, USA
- Eric Raidl
Universität Konstanz, DE
- Keith Stenning
University of Edinburgh, UK
- Jakub Szymanik
University of Amsterdam, NL
- Hans Rott
Universität Regensburg, DE
- Niels Skovgaard Olsen
Universität Göttingen, DE
- Ivan José Varzinczak
Artois University, FR
- Emil Weydert
University of Luxembourg, LU



Report from Dagstuhl Seminar 19041

New Horizons in Parameterized Complexity

Edited by

Fedor V. Fomin¹, Dániel Marx², Saket Saurabh³, and Meirav Zehavi⁴

1 University of Bergen, NO, fomin@ii.uib.no

2 Hungarian Academy of Sciences – Budapest, HU, dmarx@cs.bme.hu

3 Institute of Mathematical Sciences – Chennai, IN, saket@imsc.res.in

4 Ben Gurion University – Beer Sheva, IL, meiravze@bgu.ac.il

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19041 “New Horizons in Parameterized Complexity”.

Parameterized Complexity is celebrating its 30th birthday in 2019. In these three decades, there has been tremendous progress in developing the area. The central vision of Parameterized Complexity through all these years has been to provide the algorithmic and complexity-theoretic toolkit for studying multivariate algorithmics in different disciplines and subfields of Computer Science. These tools are universal as they did not only help in the development of the core of Parameterized Complexity, but also led to its success in other subfields of Computer Science such as Approximation Algorithms, Computational Social Choice, Computational Geometry, problems solvable in P (polynomial time), to name a few.

In the last few years, we have witnessed several exciting developments of new parameterized techniques and tools in the following subfields of Computer Science and Optimization: Mathematical Programming, Computational Linear Algebra, Computational Counting, Derandomization, and Approximation Algorithms. The main objective of the seminar was to initiate the discussion on which of the recent domain-specific algorithms and complexity advances can become useful in other domains.

Seminar January 20–25, 2019 – <http://www.dagstuhl.de/19041>

2012 ACM Subject Classification Theory of computation → Parameterized complexity and exact algorithms, Theory of computation → Fixed parameter tractability, Theory of computation → W hierarchy

Keywords and phrases Intractability, Parameterized Complexity

Digital Object Identifier 10.4230/DagRep.9.1.67

Edited in cooperation with Roohani Sharma

1 Executive Summary

Fedor V. Fomin (University of Bergen, NO)

Dániel Marx (Hungarian Academy of Sciences, HU)

Saket Saurabh (Institute of Mathematical Sciences, IN)

Meirav Zehavi (Ben Gurion University, IL)

License © Creative Commons BY 3.0 Unported license
© Fedor V. Fomin, Dániel Marx, Saket Saurabh and Meirav Zehavi

In 2019 the parameterized complexity (PC) community is celebrating two round dates: 30 years since the appearance of the paper of Abrahamson, Ellis, Fellows, and Mata in FOCS 1989, which can be considered as the starting point of PC, and 20 years since the appearance of the influential book of Downey and Fellows “*Parameterized Complexity*”.



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

New Horizons in Parameterized Complexity, *Dagstuhl Reports*, Vol. 9, Issue 1, pp. 67–87

Editors: Fedor V. Fomin, Dániel Marx, Saket Saurabh, and Meirav Zehavi



DAGSTUHL
REPORTS

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

In these three decades, there has been tremendous progress in developing the area. The central vision of Parameterized Complexity through all these years has been to provide the algorithmic and complexity-theoretic toolkit for studying multivariate algorithmics in different disciplines and subfields of Computer Science. To achieve this vision, several algorithmic and complexity theoretic tools such as polynomial time preprocessing, aka kernelization, color-coding, graph-decompositions, parameterized integer programming, iterative compression, or lower bounds methods based on assumptions stronger than $P=NP$ have been developed. These tools are universal as they did not only help in the development of the core of Parameterized Complexity, but also led to its success in other subfields of Computer Science such as Approximation Algorithms, Computational Social Choice, Computational Geometry, problems solvable in P (polynomial time) to name a few.

All cross-discipline developments result in flow of ideas and methods in both directions. In the last few years, we have witnessed several exciting developments of new parameterized techniques and tools in the following subfields of Computer Science and Optimization: Mathematical Programming, Computational Linear Algebra, Computational Counting, Derandomization, and Approximation Algorithms. A natural question is whether these domain-centric methods and tools are universal. That is, can they permeate boundaries of subfields and be employed wherever Parameterized Complexity approach can be used? The main objective of the seminar was to initiate the discussion on which of the recent domain-specific algorithms and complexity advances can become useful in other domains.

The seminar collected 46 participants from 18 countries. The participants presented their recent results in 26 invited and contributed talks. Open problems were discussed in open problem and discussion sessions.

2 Table of Contents

Executive Summary

Fedor V. Fomin, Dániel Marx, Saket Saurabh and Meirav Zehavi 67

Overview of Talks

| | |
|---|----|
| Polynomial Kernel for Interval Vertex Deletion <i>Akanksha Agrawal</i> | 71 |
| FPT inspired Approximation Algorithms <i>Henning Fernau</i> | 71 |
| On the Parameterized Complexity of Graph Modification to First-Order Logic Properties <i>Petr A. Golovach</i> | 71 |
| Parameterized Resiliency Problems via ILP <i>Gregory Gutin</i> | 72 |
| 0/1/all CSPs, Half-Integral A-path Packing, and Linear-Time FPT Algorithms <i>Yoichi Iwata</i> | 72 |
| Computing the Chromatic Number Using Graph Decompositions via Matrix Rank <i>Bart Jansen</i> | 73 |
| N-fold IP: FPT algorithm and applications <i>Martin Koutecký</i> | 74 |
| Parameterized Inapproximability: A (Semi-)Survey <i>Pasin Manurangsi</i> | 74 |
| Decompositions of Unit Disk Graphs and Algorithmic Applications <i>Meirav Zehavi</i> | 75 |
| Hitting Long Directed Cycles is Fixed-Parameter Tractable <i>Matthias Mnich</i> | 75 |
| New Algorithms for Planar Subgraph Isomorphism <i>Jesper Nederlof</i> | 75 |
| Integer Programming in Parameter-Tractable Strongly-Polynomial Time <i>Shmuel Onn</i> | 76 |
| Approximation Schemes for Low-Rank Binary Matrix Approximation Problems <i>Fahad Panolan</i> | 76 |
| On Subexponential Parameterized Algorithms for Steiner Tree and Directed Subset TSP on Planar Graphs <i>Marcin Pilipczuk</i> | 77 |
| Hitting minors on bounded treewidth graphs <i>Ignasi Sau Valls</i> | 77 |
| On a polynomial kernel for Directed Feedback Vertex Set <i>Roohani Sharma</i> | 78 |

Open Problems

| | |
|---|----|
| Shortest Three Disjoint Path <i>Andreas Björklund</i> | 79 |
| Counting forests with few components <i>Mark Jerrum</i> | 79 |
| Makespan Minimization on Identical Machines ($P C_{\max}$) by #job types <i>Martin Koutecký</i> | 80 |
| Stochastic bounding box <i>Sergio Cabello</i> | 80 |
| Tight bound for the number of multibudgeted important separators <i>Marcin Pilipczuk</i> | 81 |
| Three disjoint paths that are each shortest paths <i>Marcin Pilipczuk</i> | 82 |
| k -exchange TSP parameterized by $k + d$ <i>Yoichi Iwata</i> | 82 |
| Count k -Walks <i>Holger Dell</i> | 83 |
| Shortest Vector Problem (SVP) in ℓ_1 Norm <i>Pasin Manurangsi</i> | 83 |
| Polynomial kernel for Bicolored P_3 -Deletion <i>Christian Komusiewicz</i> | 84 |
| Metric TSP with Deadlines <i>Matthias Mnich</i> | 84 |
| Optimization over Degree Sequences <i>Shmuel Onn</i> | 85 |
| Dynamic Cluster Editing <i>Rolf Niedermeier</i> | 85 |
| Resolution <i>Stefan Szeider</i> | 86 |
| Participants | 87 |

3 Overview of Talks

3.1 Polynomial Kernel for Interval Vertex Deletion

Akanksha Agrawal (*Hungarian Academy of Sciences – Budapest, HU*)

License © Creative Commons BY 3.0 Unported license
© Akanksha Agrawal

Joint work of Akanksha Agrawal, Daniel Lokshantov, Pranabendu Misra, Saket Saurabh, Meirav Zehavi

Given a graph G and an integer k , the Interval Vertex Deletion (IVD) problem asks whether there exists a vertex subset S of size at most k , such that $G-S$ is an interval graph. The existence of a polynomial kernel for IVD remained a well-known open problem in Parameterized Complexity. In this talk we look at a sketch of polynomial kernel for the problem (with the parameter being the solution size). Over the course of talk, we will mainly focus on a kernel for IVD, when parameterized by the vertex cover number. The ideas in discussed in the above kernel is one of the key ingredients in our kernel for IVD, when parameterized by the solution size.

3.2 FPT inspired Approximation Algorithms

Henning Fernau (*Universität Trier, DE*)

License © Creative Commons BY 3.0 Unported license
© Henning Fernau

Approximation algorithms predate parameterized algorithms by quite some time. Therefore, several algorithmic ideas have been transferred from approximation to FPT. However, there are also opportunities to translate typical FPT ideas into algorithmic ideas for approximation. We will showcase this by looking at data reductions. One of the nice features that come with using approximative data reductions is that the approximation algorithm can monitor itself during execution, thereby proving that the actual approximation ratio is (possibly far) better than the typical worst-case analysis would show. We will also present experimental results that prove that this approach could work very well in practice.

References

- 1 Henning Fernau. *FPT inspired approximations*. To appear in the proceeding of “Aspects of Computation”, World Scientific, 2019.

3.3 On the Parameterized Complexity of Graph Modification to First-Order Logic Properties

Petr A. Golovach (*University of Bergen, NO*)

License © Creative Commons BY 3.0 Unported license
© Petr A. Golovach

Joint work of Petr A. Golovach, Fedor V. Fomin, Dimitrios M. Thilikos

Main reference Fedor V. Fomin, Petr A. Golovach, Dimitrios M. Thilikos: “On the Parameterized Complexity of Graph Modification to First-Order Logic Properties”, CoRR, Vol. abs/1805.04375, 2018.

URL <https://arxiv.org/abs/1805.04375>

We establish new connections between parameterized/kernelization complexity of graph modification problems and expressibility in logic. For a first-order logic formula φ , we consider

the problem of deciding whether an input graph can be modified by removing/adding at most k vertices/edges such that the resulting modification has the property expressible by φ . We provide sufficient and necessary conditions on the structure of the prefix of φ specifying when the corresponding graph modification problem is fixed-parameter tractable (parameterized by k) and when it admits a polynomial kernel.

3.4 Parameterized Resiliency Problems via ILP

Gregory Gutin (Royal Holloway, University of London, GB)

License © Creative Commons BY 3.0 Unported license

© Gregory Gutin

Joint work of Jason Crampton, Gregory Gutin, Martin Koutecký, Remi Watrigant

Main reference Jason Crampton, Gregory Z. Gutin, Rémi Watrigant: “An Approach to Parameterized Resiliency Problems Using Integer Linear Programming”, CoRR, Vol. abs/1605.08738, 2016.

URL <https://arxiv.org/abs/1605.08738>

We introduce an extension of decision problems called resiliency problems. In resiliency problems, the goal is to decide whether an instance remains positive after any (appropriately defined) perturbation has been applied to it. To tackle these kinds of problems, some of which might be of practical interest, we introduce a notion of resiliency for Integer Linear Programs (ILP) and show how to use a result of Eisenbrand and Shmonin (Math. Oper. Res., 2008) on Parametric Linear Programming to prove that ILP Resiliency is fixed-parameter tractable (FPT) under a certain parameterization.

To demonstrate the utility of our result, we consider natural resiliency versions of several concrete problems, and prove that they are FPT under natural parameterizations. Our first results concern a four-variate problem which generalizes the Disjoint Set Cover problem and which is of interest in access control. We obtain a complete parameterized complexity classification for every possible combination of the parameters. Then, we introduce and study a resiliency version of the Closest String problem, for which we extend an FPT result of Gramm et al. (Algorithmica, 2003). We also consider problems in the fields of scheduling and social choice. We believe that many other problems can be tackled by our framework.

3.5 0/1/all CSPs, Half-Integral A-path Packing, and Linear-Time FPT Algorithms

Yoichi Iwata (National Institute of Informatics – Tokyo, JP)

License © Creative Commons BY 3.0 Unported license

© Yoichi Iwata

Joint work of Yoichi Iwata, Yutaro Yamaguchi, Yuichi Yoshida

Main reference Yoichi Iwata, Yutaro Yamaguchi, Yuichi Yoshida: “0/1/All CSPs, Half-Integral A-Path Packing, and Linear-Time FPT Algorithms”, in Proc. of the 59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018, pp. 462–473, IEEE Computer Society, 2018.

URL <https://doi.org/10.1109/FOCS.2018.00051>

0/1/all CSPs can be solved in linear time by a simple DFS called a unit propagation. We consider an optimization variant of the CSPs where the objective is to delete the minimum subset of variables to make the given instance satisfiable. When the instance is unsatisfiable, the unit propagation finds a walk leading to a contradiction, and the size of the maximum half-integral packing of such walks gives a lower bound on the solution size. We provide an

$O(km)$ -time algorithm for computing the maximum half-integral packing, where k is the size of the packing and m is the number of constraints, and we show that a branch-and-bound method using this lower bound can solve the problem in linear FPT time. We also discuss several other applications.

3.6 Computing the Chromatic Number Using Graph Decompositions via Matrix Rank

Bart Jansen (TU Eindhoven, NL)

License  Creative Commons BY 3.0 Unported license

© Bart Jansen

Joint work of Bart M. P. Jansen, Jesper Nederlof

Main reference Bart M. P. Jansen, Jesper Nederlof: “Computing the Chromatic Number Using Graph Decompositions via Matrix Rank”, in Proc. of the 26th Annual European Symposium on Algorithms, ESA 2018, August 20-22, 2018, Helsinki, Finland, LIPIcs, Vol. 112, pp. 47:1–47:15, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.

URL <http://dx.doi.org/10.4230/LIPIcs.ESA.2018.47>

Computing the smallest number q such that the vertices of a given graph can be properly q -colored is one of the oldest and most fundamental problems in combinatorial optimization. The q -Coloring problem has been studied intensively using the framework of parameterized algorithms, resulting in a very good understanding of the best-possible algorithms for several parameterizations based on the structure of the graph. For example, algorithms are known to solve the problem on graphs of treewidth tw in time $O^*(q^{tw})$, while a running time of $O^*((q - \epsilon)^{tw})$ is impossible assuming the Strong Exponential Time Hypothesis (SETH). While there is an abundance of work for parameterizations based on decompositions of the graph by vertex separators, almost nothing is known about parameterizations based on edge separators. We fill this gap by studying q -Coloring parameterized by cutwidth, and parameterized by pathwidth in bounded-degree graphs. Our research uncovers interesting new ways to exploit small edge separators.

We present two algorithms for q -Coloring parameterized by cutwidth $cutw$: a deterministic one that runs in time $O^*(2^{\omega \cdot cutw})$, where ω is the matrix multiplication constant, and a randomized one with runtime $O^*(2^{cutw})$. In sharp contrast to earlier work, the running time is independent of q . The dependence on cutwidth is optimal: we prove that even 3-Coloring cannot be solved in $O^*((2 - \epsilon)^{cutw})$ time assuming SETH. Our algorithms rely on a new rank bound for a matrix that describes compatible colorings. Combined with a simple communication protocol for evaluating a product of two polynomials, this also yields an $O^*((\lfloor d/2 \rfloor + 1)^{pw})$ time randomized algorithm for q -Coloring on graphs of pathwidth pw and maximum degree d . Such a runtime was first obtained by Bjorklund, but only for graphs with few proper colorings. We also prove that this result is optimal in the sense that no $O^*((\lfloor d/2 \rfloor + 1 - \epsilon)^{pw})$ -time algorithm exists assuming SETH.

3.7 N-fold IP: FPT algorithm and applications

Martin Koutecký (Technion – Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Martin Koutecký

Joint work of Dušan Knop, Martin Koutecký, Asaf Levin, Shmuel Onn

Main reference Martin Koutecký, Asaf Levin, Shmuel Onn: “A Parameterized Strongly Polynomial Algorithm for Block Structured Integer Programs”, in Proc. of the 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic, LIPIcs, Vol. 107, pp. 85:1–85:14, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.

URL <http://dx.doi.org/10.4230/LIPIcs.ICALP.2018.85>

Integer Linear Programming is a fundamental optimization problem. Basic FPT results about ILP have been shown in the 80’s by Papadimitriou and Lenstra, and Lenstra’s algorithm has been applied extensively in parameterized complexity since 2003. A new class of IPs of variable dimension called n -fold IPs has been extensively studied since the 2000’s, culminating in an FPT algorithm in 2013, which has been used in parameterized complexity for the first time in 2016. Since then, several important applications as well as extensions and improvements of this algorithm have been found.

In this talk I will define n -fold IPs, briefly overview the FPT algorithm solving it, and then focus on two classes of application. The first class concerns CLOSEST STRING-type problems and BRIBERY-type problems, for which the application of n -fold IP has led to the first single-exponential algorithms. The second class concerns problems from scheduling, where n -fold IP is the only known technique which yields FPT results for several fundamental problems such as minimization of sum of weighted completion times, or makespan minimization when machines have many different speeds. Finally, I will point out what I believe to be the two most important open problems in the area.

3.8 Parameterized Inapproximability: A (Semi-)Survey

Pasin Manurangsi (University of California – Berkeley, US)

License  Creative Commons BY 3.0 Unported license
© Pasin Manurangsi

Joint work of Parinya Chalermsook, Marek Cygan, Karthik C. S., Guy Kortsarz, Bundit Laekhanukit, Pasin Manurangsi, Danupon Nanongkai, Luca Trevisan

Main reference Karthik C. S., Bundit Laekhanukit, Pasin Manurangsi: “On the parameterized complexity of approximating dominating set”, in Proc. of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, pp. 1283–1296, ACM, 2018.

URL <https://doi.org/10.1145/3188745.3188896>

Main reference Parinya Chalermsook, Marek Cygan, Guy Kortsarz, Bundit Laekhanukit, Pasin Manurangsi, Danupon Nanongkai, Luca Trevisan: “From Gap-ETH to FPT-Inapproximability: Clique, Dominating Set, and More”, in Proc. of the 58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017, pp. 743–754, IEEE Computer Society, 2017.

URL <https://doi.org/10.1109/FOCS.2017.74>

In this talk, I will survey some of the recent results on parameterized inapproximability, with focus on the total inapproximability of k -Dominating Set and k -Clique.

3.9 Decompositions of Unit Disk Graphs and Algorithmic Applications

Meirav Zehavi (Ben Gurion University – Beer Sheva, IL)

License © Creative Commons BY 3.0 Unported license
© Meirav Zehavi

Joint work of Fedor Fomin, Daniel Lokshtanov, Fahad Panolan, Saket Saurabh, Meirav Zehavi

Main reference Fedor V. Fomin, Daniel Lokshtanov, Saket Saurabh: “Excluded Grid Minors and Efficient Polynomial-Time Approximation Schemes”, *J. ACM*, Vol. 65(2), pp. 10:1–10:44, 2018.

URL <https://doi.org/10.1145/3154833>

Main reference Fedor V. Fomin, Danile Lokshtanov, Fahad Panolan, Saket Saurabh, Meirav Zehavi: “Finding, Hitting and Packing Cycles in Subexponential Time on Unit Disk Graphs”, in *Discrete and Computational Geometry*, 2019.

URL <https://doi.org/10.1007/s00454-018-00054-x>

Main reference Fahad Panolan, Saket Saurabh, Meirav Zehavi: “Contraction Decomposition in Unit Disk Graphs and Algorithmic Applications in Parameterized Complexity”, in *Proc. of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pp. 1035–1054, SIAM, 2019.

URL <https://doi.org/10.1137/1.9781611975482.64>

In this talk, I will discuss decompositions of unit disk graphs with applications in the design of subexponential and exponential time parameterized algorithms.

3.10 Hitting Long Directed Cycles is Fixed-Parameter Tractable

Matthias Mnich (Universität Bonn, DE)

License © Creative Commons BY 3.0 Unported license
© Matthias Mnich

Joint work of Alexander Göke, Dániel Marx, Matthias Mnich

The Directed Feedback Vertex Set (DFVS) problem takes as input a directed graph G and seeks a minimum-size vertex set S that hits all cycles in G ; this is one of Karp’s 21 NP-complete problems. Resolving the parameterized complexity status of the DFVS problem was a long-standing open problem until Chen et al. (STOC 2008, J.ACM 2008) showed its fixed-parameter tractability via a $4^k k! n^{O(1)}$ -time algorithm, where $k = |S|$. We give consider the wide generalization of the DFVS problem where we want to intersect/long/ directed cycles: find a minimum-size set S of arcs or vertices such that every simple directed cycle of $G - S$ has length at most ℓ . Our main result is an algorithm which solves this problem in time $2^{O(\ell^{3k^3 \log k + k^6 \log k \log \ell})} n^{O(1)}$. Our algorithm therefore provides an exact version of the Erdős-Pósa property for long cycles in directed graphs, which was recently proved by Kreutzer and Kawarabayashi [STOC 2015].

3.11 New Algorithms for Planar Subgraph Isomorphism

Jesper Nederlof (TU Eindhoven, NL)

License © Creative Commons BY 3.0 Unported license
© Jesper Nederlof

Joint work of Jesper Nederlof

Main reference Jesper Nederlof: “Detecting and Counting Small Patterns in Planar Graphs”, CoRR, arXiv:1904.11285v1 [cs.DS], 2019.

URL <https://arxiv.org/abs/1904.11285v1>

We present sub exponential time algorithms for finding and counting (induced) patterns in planar graphs.

3.12 Integer Programming in Parameter-Tractable Strongly-Polynomial Time

Shmuel Onn (Technion – Haifa, IL)

License © Creative Commons BY 3.0 Unported license
© Shmuel Onn

Joint work of Martin Koutecký, Asaf Levin, Shmuel Onn

Main reference Martin Koutecký, Asaf Levin, Shmuel Onn: “A Parameterized Strongly Polynomial Algorithm for Block Structured Integer Programs”, in Proc. of the 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic, LIPIcs, Vol. 107, pp. 85:1–85:14, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.

URL <http://dx.doi.org/10.4230/LIPIcs.ICALP.2018.85>

Integer programming has been a powerful tool in classical combinatorial optimization due to its broad modeling power.

We establish a new fundamental FPT result on integer programming and hope it will provide a new tool that may allow to establish new FPT results for a variety of combinatorial optimization problems. We will be happy to learn of any such progress that may occur.

The result, which extends, improves, unifies and simplifies many results of the last decade, is the following.

Theorem: Integer programming can be solved in fixed parameter-tractable strongly-polynomial time $f(a,d)\text{poly}(n)$, for some polynomial of the number n of variables, and some function f of the maximum absolute value a of any entry of the matrix A defining the integer program and the minimum d between the treedepth of A and the treedepth of its transpose.

The slides of the talk are available below and on my homepage.

3.13 Approximation Schemes for Low-Rank Binary Matrix Approximation Problems

Fahad Panolan (University of Bergen, NO)

License © Creative Commons BY 3.0 Unported license
© Fahad Panolan

Joint work of Fedor V. Fomin, Petr A. Golovach, Daniel Lokshtanov, Fahad Panolan, Saket Saurabh

Main reference Fedor V. Fomin, Petr A. Golovach, Daniel Lokshtanov, Fahad Panolan, Saket Saurabh: “Approximation Schemes for Low-Rank Binary Matrix Approximation Problems”, CoRR, Vol. abs/1807.07156, 2018.

URL <https://arxiv.org/abs/1807.07156>

We provide a randomized linear time approximation scheme for a generic problem about clustering of binary vectors subject to additional constraints. The new constrained clustering problem encompasses a number of problems and by solving it, we obtain the first linear time-approximation schemes for a number of well-studied fundamental problems concerning clustering of binary vectors and low-rank approximation of binary matrices. Our algorithm runs in time $f(k, \epsilon) \cdot n \cdot m$, where f is some computable function, k is the number of clusters, n is the number of binary vectors in the input and m is the dimension of these vectors.

3.14 On Subexponential Parameterized Algorithms for Steiner Tree and Directed Subset TSP on Planar Graphs

Marcin Pilipczuk (University of Warsaw, PL)

License © Creative Commons BY 3.0 Unported license
© Marcin Pilipczuk

Joint work of Dániel Marx, Marcin Pilipczuk, Michał Pilipczuk

Main reference Dániel Marx, Marcin Pilipczuk, Michał Pilipczuk: “On Subexponential Parameterized Algorithms for Steiner Tree and Directed Subset TSP on Planar Graphs”, in Proc. of the 59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018, pp. 474–484, IEEE Computer Society, 2018.

URL <https://doi.org/10.1109/FOCS.2018.00052>

There are numerous examples of the so-called “square root phenomenon” in the field of parameterized algorithms: many of the most fundamental graph problems, parameterized by some natural parameter k , become significantly simpler when restricted to planar graphs and in particular the best possible running time is exponential in $\mathcal{O}(\sqrt{k})$ instead of $\mathcal{O}(k)$ (modulo standard complexity assumptions). We consider two classic optimization problems parameterized by the number of terminals. The STEINER TREE problem asks for a minimum-weight subtree connecting a given set of terminals T in an edge-weighted graph. In the SUBSET TRAVELING SALESMAN problem we are asked to visit all the terminals T by a minimum-weight closed walk. We investigate the parameterized complexity of these problems in planar graphs, where the number $k = |T|$ of terminals is regarded as the parameter. Our results are the following:

- SUBSET TSP can be solved in time $2^{\mathcal{O}(\sqrt{k} \log k)} \cdot n^{\mathcal{O}(1)}$ even on edge-weighted directed planar graphs. This improves upon the algorithm of Klein and Marx [SODA 2014] with the same running time that worked only on undirected planar graphs with polynomially large integer weights.
- Assuming the Exponential-Time Hypothesis, STEINER TREE on undirected planar graphs cannot be solved in time $2^{o(k)} \cdot n^{\mathcal{O}(1)}$, even in the unit-weight setting. This lower bound makes STEINER TREE the first “genuinely planar” problem (i.e., where the input is only planar graph with a set of distinguished terminals) for which we can show that the square root phenomenon does not appear.
- STEINER TREE can be solved in time $n^{\mathcal{O}(\sqrt{k})} \cdot W$ on undirected planar graphs with maximum edge weight W . Note that this result is incomparable to the fact that the problem is known to be solvable in time $2^k \cdot n^{\mathcal{O}(1)}$ even in general graphs.

A direct corollary of the combination of our results for STEINER TREE is that this problem does not admit a parameter-preserving polynomial kernel on planar graphs unless ETH fails.

3.15 Hitting minors on bounded treewidth graphs

Ignasi Sau Valls (CNRS – Montpellier, FR)

License © Creative Commons BY 3.0 Unported license
© Ignasi Sau Valls

Joint work of Julien Baste, Ignasi Sau, Dimitrios M. Thilikos

Main reference Julien Baste, Ignasi Sau, Dimitrios M. Thilikos: “Hitting minors on bounded treewidth graphs”, CoRR, Vol. abs/1704.07284, 2017.

URL <https://arxiv.org/abs/1704.07284>

For a fixed collection of graphs \mathcal{F} , the \mathcal{F} -M-Deletion problem consists in, given a graph G and an integer k , decide whether there exists $S \subseteq V(G)$ with $|S| \leq k$ such that $G \setminus S$

does not contain any of the graphs in F as a minor. We are interested in its parameterized complexity when the parameter is the treewidth of G , denoted by tw . Our objective is to determine, for a fixed F , the smallest function f_F such that F-M-Deletion can be solved in time $f_F(tw) \cdot n^{O(1)}$ on n -vertex graphs. We prove that $f_F(tw) = 2^{2^{O(tw \cdot \log tw)}}$ for every collection F , that $f_F(tw) = 2^{O(tw \cdot \log tw)}$ if all the graphs in F are connected and at least one of them is planar, and that $f_F(tw) = 2^{O(tw)}$ if in addition the input graph G is planar or embedded in a surface. When F contains a single connected planar graph H , we obtain a tight dichotomy about the asymptotic complexity of H-M-Deletion. Namely, we prove that $f_H(tw) = 2^{\theta(tw)}$ if H is a minor of the banner (that is, the graph consisting of a C_4 plus a pendent edge) that is different from P_5 , and that $f_H(tw) = 2^{\theta(tw \cdot \log tw)}$ otherwise. All the lower bounds hold under the ETH. We also consider the version of the problem where the graphs in F are forbidden as topological minors, and prove similar results, except that, in the algorithms, instead of requiring F to contain a planar graph, we need it to contain a subcubic planar graph. We also prove that, for this problem, $f_{K_{1,i}}(tw) = 2^{\theta(tw)}$ for every $i \geq 1$, while for the minor version it holds that $f_{K_{1,i}}(tw) = 2^{\theta(tw \cdot \log tw)}$ for every $i \geq 4$.

3.16 On a polynomial kernel for Directed Feedback Vertex Set

Roohani Sharma (*Institute of Mathematical Sciences – Chennai, IN*)

License  Creative Commons BY 3.0 Unported license
© Roohani Sharma

Joint work of Daniel Lokshtanov, M. S. Ramanujan, Saket Saurabh, Roohani Sharma, Meirav Zehavi

Main reference Daniel Lokshtanov, M. S. Ramanujan, Saket Saurabh, Roohani Sharma, Meirav Zehavi: “Brief Announcement: Treewidth Modulator: Emergency Exit for DFVS”, in Proc. of the 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic, LIPIcs, Vol. 107, pp. 110:1–110:4, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.

URL <http://dx.doi.org/10.4230/LIPIcs.ICALP.2018.110>

In the Directed Feedback Vertex Set (DFVS), the input is a directed graph D and an integer k , and the question is to determine whether there exists a set of vertices of D of size at most k whose removal makes the digraph acyclic. The problem concerning the existence of a polynomial kernel for DFVS is an interesting and challenging open problem in the field of parameterized complexity. We take a step towards answering this question by giving a polynomial kernel for DFVS with an enriched parameter. In particular, we study DFVS parameterized by the solution size (k) and the size of a treewidth- η modulator of the underlying undirected graph of D (say ℓ). In particular, we give a kernel for DFVS of size $(k + \ell)^{O(1)}$. This result also generalizes the result by Bergougnoux et al. that gives a polynomial kernel for DFVS parameterized by the feedback vertex set of the underlying undirected graph of D . As a corollary, our result implies a polynomial kernel for DFVS on instances that are $k^{O(1)}$ vertices away from having bounded treewidth.

4 Open Problems

4.1 Shortest Three Disjoint Path

Andreas Björklund (Lund University, SE)

License  Creative Commons BY 3.0 Unported license
© Andreas Björklund

Decide the complexity of the Shortest Three Disjoint Paths problem: Given an undirected unweighted graph $G = (V, E)$ and three pairs of distinct terminal vertices (s_1, t_1) , (s_2, t_2) , and $(s_3, t_3) \in V \times V$, find three pairwise vertex disjoint paths connecting s_i with t_i for $i = 1, 2, 3$ in G , respectively, of minimum total length (the number of edges in the three paths). Already the restriction to planar graphs of maximum degree three is open. For Shortest *Two* Disjoint Paths, a randomized polynomial time algorithm is known [1], and in the planar maximum degree three case there is a deterministic polynomial time algorithm that also counts the number of solutions [2]. For Shortest Three Disjoint Paths in planar graphs a deterministic polynomial time algorithm is known when all terminals lie on the same face [3].

References

- 1 Andreas Björklund and Thore Husfeldt. Shortest two disjoint paths in polynomial time. In *International Colloquium on Automata, Languages, and Programming*, pages 211–222. Springer, 2014.
- 2 Andreas Björklund and Thore Husfeldt. Counting shortest two disjoint paths in cubic planar graphs with an nc algorithm. *arXiv preprint arXiv:1806.07586*, 2018.
- 3 Samir Datta, Siddharth Iyer, Raghav Kulkarni, and Anish Mukherjee. Shortest k -disjoint paths via determinants. *arXiv preprint arXiv:1802.01338*, 2018.

4.2 Counting forests with few components

Mark Jerrum (Queen Mary University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Mark Jerrum

Suppose G is an undirected graph on n vertices. Kirchhoff’s Matrix-tree Theorem expresses the number of spanning trees in G as the determinant of an $(n - 1) \times (n - 1)$ matrix, thus providing a polynomial-time algorithm for counting (exactly) the spanning trees in G . Building on this, Liu and Chow [2] gave a method to count the number of $(k + 1)$ -component (spanning) forests in G , for any $k \geq 0$. Their recursive procedure is polynomial-time for any fixed k , but the exponent grows with k . In modern terminology, they showed that counting $(k + 1)$ -component forests is in XP. My question is **whether counting $(k + 1)$ -component forests is in FPT**.

The general form of this open problem is: Start with a set of structures having some property Π , for example, Π might be the property of being a spanning tree of a graph G . Assume that there is a polynomial-time algorithm for counting structures with property Π . Now perturb the property Π to Π' , and consider the derived problem of counting structures with property Π' . Introduce a parameter k to measure the extent of the perturbation, for example, k might be the number of “missing edges” in a spanning tree. It is natural to ask whether this perturbed counting problem is in FPT or XP, or is $\#W[1]$ -hard, etc., regarded

as a problem parameterized by k . An example of a solved problem of this type is counting matchings in a planar graph with $2k$ uncovered vertices or “monomers”, which was shown to be $\#W[1]$ -hard by Curticapean [1].

Returning to spanning trees, it is natural to perturb the structures in the opposite direction and consider trees with “excess edges”, in other words, connected spanning subgraphs with $n + k - 1$ edges. (The parameterization is chosen so that spanning trees correspond again to $k = 0$). Surprisingly, despite the obvious similarity to counting forests (indeed the problems are dual in matroid theoretic terms), it is not even known whether the counting problem is in XP. Note that there is no contradiction here: the class of graphic matroids is not closed under taking duals.

References

- 1 Radu Curticapean. Counting matchings with k unmatched vertices in planar graphs. In *24th Annual European Symposium on Algorithms*, volume 57 of *LIPICs. Leibniz Int. Proc. Inform.*, pages Art. No. 33, 17. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.
- 2 C. J. Liu and Yutze Chow. Enumeration of forests in a graph. *Proc. Amer. Math. Soc.*, 83(3):659–662, 1981.

4.3 Makespan Minimization on Identical Machines ($P||C_{\max}$) by #job types

Martin Koutecký (Technion, Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Martin Koutecký

In the Makespan Minimization on Identical Machines ($P||C_{\max}$) by #job types, we are given m identical machines, τ types of jobs, $n_j \in \mathbb{N}$ jobs of type $j \in [\tau]$, where each job of type $j \in [\tau]$ has processing time $p_j \in \mathbb{N}$, and the question is to find a schedule minimizing the makespan C_{\max} , i.e., the time when the last job finishes.

Goemans and Rothvoss [1] have shown that the problem is solvable in time roughly $\mathcal{O}^*((\log p_{\max})^{2^\tau})$, where $p_{\max} = \max_j p_j$. With respect to τ this is an XP algorithm, or an FPT algorithm if p_{\max} has size polynomial in the encoding length of the instance. It remains open **whether the problem is FPT or $W[1]$ -hard**.

References

- 1 Michel X Goemans and Thomas Rothvoß. Polynomiality for bin packing with a constant number of item types. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 830–839. Society for Industrial and Applied Mathematics, 2014.

4.4 Stochastic bounding box

Sergio Cabello (University of Ljubljana, SI)

License  Creative Commons BY 3.0 Unported license
© Sergio Cabello

Let P be a set of n points in \mathbb{R}^d and assume that each point p of P has a number $\pi(p) \in (0, 1]$ associated to it. We construct a random subset R of P where we include each point p of P with probability $\pi(p)$, where the decision for each point is made independently. We want

to compute the expected volume of the minimum axis-parallel box that contains R . In the plane this can be done in $O(n \log n)$ time [1], assuming that each arithmetic operation takes constant time. Using the 2-dimensional case as base case, one can solve the problem in $O(n^{d-1} \log n)$ time for each $d \geq 3$. The non-stochastic version can be solved in $O(dn)$ trivially. **Is the problem $W[1]$ -hard or FPT when parameterized by the dimension d ?**

References

- 1 Constantinos Tsirogiannis, Frank Staals, and Vincent Pellissier. Computing the expected value and variance of geometric measures. *Journal of Experimental Algorithmics (JEA)*, 23(2):2–4, 2018.

4.5 Tight bound for the number of multibudgeted important separators

Marcin Pilipczuk (University of Warsaw, PL)

License  Creative Commons BY 3.0 Unported license
© Marcin Pilipczuk

Consider a directed graph G with distinguished source $s \in V(G)$, sink $t \in V(G)$, and a partition $E(G) = E_1 \uplus E_2 \uplus \dots \uplus E_\ell$ of the arc set. A set $C \subseteq E(G)$ is an $s-t$ separator if there is no path from s to t in $G - C$. An $s-t$ separator C is a *minimal $s-t$ separator* if no proper subset of C is an $s-t$ separator. A minimal $s-t$ separator D *dominates* a minimal $s-t$ separator C if every vertex reachable from s in $G - C$ is also reachable from s in $G - D$ and for every $i \in [\ell]$ we have $|C \cap E_i| \geq |D \cap E_i|$. A minimal $s-t$ separator D is *important* if no other minimal $s-t$ separator dominates it. The classic result asserts that for the single-budget case $\ell = 1$ there are at most 4^k important separators of size at most k [3, 1].

In our recent IPEC'18 paper [2] we show a generalization of this result for multibudgeted case with a bound of $2^{\mathcal{O}(k^2 \log k)}$ for the number of multibudgeted important separators of size at most k . However, the best known lower bound is $2^{\Omega(k \log k)}$ attained via the following simple construction. Let $\ell = k$ and let G consist of k paths $(P_j)_{j=1}^k$; each path P_j starts in s , ends in t , and consists of k edges $(e_j^i)_{i=1}^k$ in this order such that $e_j^i \in E_i$ for every $i, j \in [k]$. Then any minimal $s-t$ separator in G is a multibudgeted important separator. **Please close the gap.**

References

- 1 Jianer Chen, Yang Liu, Songjian Lu, Barry O'Sullivan, and Igor Razgon. A fixed-parameter algorithm for the directed feedback vertex set problem. *J. ACM*, 55(5):21:1–21:19, 2008.
- 2 Stefan Kratsch, Shaohua Li, Dániel Marx, Marcin Pilipczuk, and Magnus Wahlström. Multi-budgeted directed cuts. *CoRR*, abs/1810.06848, 2018. Presented at IPEC 2018.
- 3 Dániel Marx. Parameterized graph separation problems. *Theor. Comput. Sci.*, 351(3):394–406, 2006.

4.6 Three disjoint paths that are each shortest paths

Marcin Pilipczuk (University of Warsaw, PL)

License  Creative Commons BY 3.0 Unported license
© Marcin Pilipczuk

Let G be an edge-weighted graph and let $(s_i, t_i)_{i=1}^k$ be terminal pairs in G . The task is to find paths $(P_i)_{i=1}^k$ such that each P_i is a shortest path in G from s_i to t_i and the paths P_i are pairwise vertex-disjoint. The problem has been introduced by Eilam-Tzoreff [2] who showed that it is polynomial-time solvable for $k = 2$ in undirected graphs with strictly positive edge weights. Bérczi and Kobayashi [1] generalized this result to $k = 2$ in directed graphs with strictly positive edge weights and later two independent groups [3, 4] showed also a generalization to $k = 2$ in undirected graphs with nonnegative edge weights. However, the **three terminal pair case remains widely open**, even in undirected unweighted graphs.

References

- 1 Kristóf Bérczi and Yusuke Kobayashi. The directed disjoint shortest paths problem. In *ESA*, volume 87 of *LIPICs*, pages 13:1–13:13. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2017.
- 2 Tali Eilam-Tzoreff. The disjoint shortest paths problem. *Discrete Applied Mathematics*, 85(2):113–138, 1998.
- 3 Marinus Gottschau, Marcus Kaiser, and Clara Waldmann. The undirected two disjoint shortest paths problem. *Oper. Res. Lett.*, 47(1):70–75, 2019.
- 4 Yusuke Kobayashi and Ryo Sako. Two disjoint shortest paths problem with non-negative edge length. *Oper. Res. Lett.*, 47(1):66–69, 2019.

4.7 k -exchange TSP parameterized by $k + d$

Yoichi Iwata (National Institute of Informatics, Tokyo, JP)

License  Creative Commons BY 3.0 Unported license
© Yoichi Iwata

In the k -exchange TSP problem, given as input an undirected graph $G = (V, E)$ with maximum degree d , a TSP tour $C \subseteq E$, and an integer k , the goal is to decide if there is a shorter TSP tour C' with $|C \setminus C'| \leq k$?

Local search with the above k -exchange neighborhoods is widely used in heuristic TSP solvers. When parameterized by k only, Marx [2] proved W[1]-hardness. In order to make the local search practical, state-of-the-art local search solvers use the following two heuristics.

1. Sparsify the graph by picking top- d important incident edges for each vertex. For example, LKH [1] uses α -nearness as the importance measure and reduces the degrees to $d = 5$.
2. Focus on sequential moves. A k -exchange move C' is called sequential if the symmetric difference $C \Delta C'$ forms a simple cycle. When the maximum degree is d , the exhaustive search for sequential k -moves runs in $d^{O(k)}n$ time.

If k -exchange TSP parameterized by $k + d$ is W[1]-hard, we can justify focusing on the sequential moves, and if it is FPT, we may have practical improvements.

References

- 1 Keld Helsgaun. An effective implementation of the lin-kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106–130, 2000.
- 2 Dániel Marx. Searching the k -change neighborhood for TSP is W[1]-hard. *Oper. Res. Lett.*, 36(1):31–36, 2008.

4.8 Count k -Walks

Holger Dell (Universität des Saarlandes, DE)

License © Creative Commons BY 3.0 Unported license
© Holger Dell

Given a graph G with n vertices and m edges, and a number k , the goal is to compute the number of all k -step walks in G . If A is the adjacency matrix of G , this number is just the sum of all entries in A^k . It can thus be computed in time $O(\log k \cdot n^\omega)$ where ω is the matrix multiplication constant. It can also be computed in time $O(k(n+m))$. The open question is: **can the problem be computed in time $o(k) \cdot (n+m)$ or even $O(n+m)$?** Or would this violate some complexity hypothesis?

4.9 Shortest Vector Problem (SVP) in ℓ_1 Norm

Pasin Manurangsi (University of California – Berkeley, USA)

License © Creative Commons BY 3.0 Unported license
© Pasin Manurangsi

Definition: Given a lattice $\mathcal{L} \subseteq \mathbb{Z}^n$ (specified by its basis), determine whether there exists a nonzero vector $\mathbf{x} \in \mathcal{L}$ whose ℓ_1 norm is at most k (i.e. $\sum_{i \in [n]} |x_i| \leq k$). Here k is our parameter. *Relevant Literature:* The non-parameterized version of the problem is very well-studied. In particular, the NP-hardness (under randomized reductions) was proved by Ajtai [1]. Later, it was shown to be hard to approximate to factor of $(2-\varepsilon)$ by Micciancio [6]. Subsequently, the factor was dramatically improved to $2^{(\log n)^{0.5-\varepsilon}}$ by Regev and Rosen [7] and then to $2^{(\log n)^{1-o(1)}}$ by Haviv and Regev [4].

The issue in adapting these proofs is that the aforementioned reductions inherently produce non-integral lattices. In particular, the lattices in [1] and [6] are irrational, whereas the lattices from [7, 4] comes from norm embeddings from ℓ_2 to ℓ_1 which, if discretize, does not result in any valuable parameter anymore.

In recent works [2, 3], it was shown that SVP in ℓ_p is $W[1]$ -hard (under randomized reductions) for all $p > 1$, but the proof fails for $p = 1$. The approach taken there was adapted from the work of Khot [5], which fails for $p = 1$ due to technical reasons. (Note that this is true even for the non-parameterized case.)

It would be nice if the **parameterized hardness of SVP** in ℓ_1 norm can be established. On the other hand, any **non-trivial positive results** would be interesting. For instance, even an $f(k)$ -FPT-approximation (for some function f) would be very nice, since this would also lead to a $g(k)$ -FPT-approximation (for some function g) for SVP in ℓ_2 norm as well.

References

- 1 Miklós Ajtai. The shortest vector problem in ℓ_2 is NP-hard for randomized reductions (extended abstract). In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 10–19, 1998.
- 2 Arnab Bhattacharyya, Suprovat Ghoshal, Karthik C. S., and Pasin Manurangsi. Parameterized intractability of even set and shortest vector problem from gap-eth. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, pages 17:1–17:15, 2018.
- 3 Édouard Bonnet, László Egri, and Dániel Marx. Fixed-parameter approximability of boolean mincsp. *CoRR*, abs/1601.04935, 2016.

- 4 Ishay Haviv and Oded Regev. Tensor-based hardness of the shortest vector problem to within almost polynomial factors. *Theory of Computing*, 8(1):513–531, 2012.
- 5 Subhash Khot. Hardness of approximating the shortest vector problem in lattices. *J. ACM*, 52(5):789–808, 2005.
- 6 Daniele Micciancio. The shortest vector in a lattice is hard to approximate to within some constant. *SIAM J. Comput.*, 30(6):2008–2035, 2000.
- 7 Oded Regev and Ricky Rosen. Lattice problems and norm embeddings. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 447–456, 2006.

4.10 Polynomial kernel for Bicolored P_3 -Deletion

Christian Komusiewicz (Universität Marburg, DE)

License  Creative Commons BY 3.0 Unported license
 © Christian Komusiewicz

In the BICOLORED P_3 -DELETION problem, given a graph $G = (V, E)$, where E is partitioned into a set E_r of red edges and a set E_b of blue edges, and an integer $k \in \mathbb{N}$, the goal is to decide whether we can delete at most k edges from G such that the remaining graph contains no bicolored P_3 as induced subgraph? Here a *bicolored P_3* is a path on three vertices with one blue and one red edge. The open question is : **Does Bicolored P_3 -Deletion parameterized by k admit a polynomial-size problem kernel?**

The reference for this is [1].

References

- 1 Niels Grüttemeier, Christian Komusiewicz, Jannik Schestag, and Frank Sommer. Destroying bicolored P_3 's by deleting few edges. *CoRR*, abs/1901.03627, 2019.

4.11 Metric TSP with Deadlines

Matthias Mnich (Universität Bonn, DE)

License  Creative Commons BY 3.0 Unported license
 © Matthias Mnich

In the METRIC TSP WITH DEADLINES problem, we are given an instance of METRIC TSP with n cities, out of which a subset of $k \ll n$ cities is distinguished. Additionally, an integer D is provided as input. The goal is to find a minimum-cost tour that visits all distinguished cities before the deadline D , or concludes that no such tour exists.

Böckenhauer et al. [1] showed that this problem admits a 2.5-approximation in time $f(k) \cdot n^{O(1)}$, and also proved a lower bound of 2 on the approximability of the problem by fixed-parameter algorithms under the assumption that $P \neq NP$. The open problem is to **close the gap between 2 and 2.5 on the fixed-parameter approximability of this problem.**

References

- 1 Hans-Joachim Böckenhauer, Juraž Hromkovic, Joachim Kneis, and Joachim Kupke. The parameterized approximability of tsp with deadlines. *Theory of Computing Systems*, 41(3):431–444, 2007.

4.12 Optimization over Degree Sequences

Shmuel Onn (Technion, Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Shmuel Onn

We consider *Optimization over Degree Sequences* problem defined below. Given functions $f_1, \dots, f_n : \{0, 1, \dots, n\} \rightarrow \mathbb{R}$, find a graph G on $[n]$ maximizing $\sum_{i=1}^n f_i(d_i(G))$, where $d_i(G)$ is the degree of vertex i in G .

We know quite a little about the complexity of this problem. If all the functions are the same $f_1 = \dots = f_n = g_1$ then we can solve it in polynomial time [2]. What is the **complexity if each f_i equals one of two given functions g_1, g_2** ? What is the **(parameterized) complexity if each f_i equals one of k given functions g_1, \dots, g_k** ? What is the **(parameterized) complexity for restricted classes of functions (e.g., convex, concave, k -piecewise linear)**?

The problem can obviously be generalized to r -uniform hypergraphs. However, then some severe restrictions on the functions should be applied, since already the following decision problem is NP-complete: given d_1, \dots, d_n , is there a 3-uniform hypergraph H with $d_i(H) = d_i$ for all i ? (The NP-completeness of this is shown in [2] solving a long open problem from [1].)

We will be happy to learn of any progress on this problem that you may come up with.

References

- 1 Charles J Colbourn, William L Kocay, and Douglas R Stinson. Some np-complete problems for hypergraph degree sequences. *Discrete applied mathematics*, 14(3):239–254, 1986.
- 2 Antoine Deza, Asaf Levin, Syed M Meesum, and Shmuel Onn. Optimization over degree sequences. *SIAM Journal on Discrete Mathematics*, 32(3):2067–2079, 2018.

4.13 Dynamic Cluster Editing

Rolf Niedermeier (TU Berlin, DE)

License  Creative Commons BY 3.0 Unported license
© Rolf Niedermeier

We consider the following “dynamic version” of a well-studied graph-based data clustering problem.

Dynamic Cluster Completion with Edge-Based Distance: In this problem the input is an undirected graph G and a cluster graph¹ G_c over the same vertex set, and two nonnegative integers, a budget k and a distance bound d . The question is to decide if there exists a cluster graph G'_c with $E(G) \subseteq E(G'_c)$ such that

- $|E(G) \oplus E(G'_c)| \leq k$ and
- $|E(G_c) \oplus E(G'_c)| \leq d$?

Herein, \oplus denotes the symmetric difference between two sets. This is a simple (still NP-hard) version of DYNAMIC CLUSTER EDITING (where adding and deleting edges from G is allowed in order to generate a cluster graph) restricted to edge additions.

¹ That is, a disjoint union of cliques.

It is open whether **Dynamic Cluster Completion with Edge-Based Distance is fixed-parameter tractable when parameterized by k** . Notably, in the conference version of Luo et al. [1] at FSTTCS 2018 DYNAMIC CLUSTER COMPLETION WITH EDGE-BASED DISTANCE was erroneously claimed to be fixed-parameter tractable for parameter k ; the proof was flawed.

DYNAMIC CLUSTER COMPLETION WITH EDGE-BASED DISTANCE is known to be fixed-parameter tractable when parameterized by d and it has a polynomial kernel when parameterized by $k + d$ [1]. Refer to Luo et al. [1] for motivation in terms of compromise clustering, local search, and target cluster graphs. Luo et al. study several variants of DYNAMIC CLUSTER EDITING, listing some further open problems.

References

- 1 Junjie Luo, Hendrik Molter, André Nichterlein, and Rolf Niedermeier. Parameterized dynamic cluster editing. *arXiv preprint arXiv:1810.06625*, 2018.

4.14 Resolution

Stefan Szeider (TU Wien, AT)

License  Creative Commons BY 3.0 Unported license
© Stefan Szeider

The *incidence graph* of a CNF formula F is the bipartite graph between the clauses and the variables of F , where a clause C and a variable x are adjacent if x appears negated or unnegated in C . The *primal graph* of F has as vertices the variables of F , two variables are adjacent if they appear together negated or unnegated in a clause of F . By *resolution* we can obtain from clauses $C \vee x$ and $D \vee \neg x$ the clause $C \vee D$. A *resolution refutation of F of size t* is a sequence C_1, \dots, C_t of clauses such that C_t is the empty clause and for each $i \in [t]$, either $C_i \in F$ or C_i can be obtained from C_j and C_ℓ for some $1 \leq j, \ell < i$. The open question is: **Do unsatisfiable CNF formulas have FPT-sized resolution refutations, parameterized by the treewidth of the incidence graph?**

Known results: (i) CNF formulas have FPT-sized resolution refutations, parameterized by the pathwidth of the incidence graph [1], and (ii) parameterized by the treewidth of the primal graph [3]. (iii) If F is a 3CNF formula, then it has an FPT-sized resolution refutation parameterized by the treewidth k of the incidence graph, since the treewidth of the primal graph is then at most $3k + 2$ [2]. (iv) One can transform in polynomial time a CNF formula F whose incidence graph has treewidth k into an equisatisfiable 3CNF formula F' whose primal graph has treewidth at most $3k + 3$ [4]; if F' is unsatisfiable it has an FPT-sized resolution refutation by (iii) but contains additional variables that were not in F .

References

- 1 Kensuke Imanishi. An upper bound for resolution size: Characterization of tractable sat instances. In *International Workshop on Algorithms and Computation*, pages 359–369. Springer, 2017.
- 2 Phokion G Kolaitis and Moshe Y Vardi. Conjunctive-query containment and constraint satisfaction. *Journal of Computer and System Sciences*, 61(2):302–332, 2000.
- 3 Irina Rish and Rina Dechter. Resolution versus search: Two strategies for sat. *Journal of Automated Reasoning*, 24(1-2):225–275, 2000.
- 4 Marko Samer and Stefan Szeider. Constraint satisfaction with bounded treewidth revisited. *Journal of Computer and System Sciences*, 76(2):103–114, 2010.

Participants

- Akanksha Agrawal
Hungarian Academy of Sciences –
Budapest, HU
- Andreas Björklund
Lund University, SE
- Sergio Cabello
University of Ljubljana, SI
- Parinya Chalermsook
Aalto University, FI
- Yijia Chen
Fudan University – Shanghai, CN
- Radu Curticapean
IT University of
Copenhagen, DK
- Holger Dell
Universität des Saarlandes, DE
- Friedrich Eisenbrand
EPFL – Lausanne, CH
- Henning Fernau
Universität Trier, DE
- Fedor V. Fomin
University of Bergen, NO
- Archontia C. Giannopoulou
TU Berlin, DE
- Petr A. Golovach
University of Bergen, NO
- Gregory Z. Gutin
Royal Holloway, University of
London, GB
- Danny Hermelin
Ben Gurion University –
Beer Sheva, IL
- Yoichi Iwata
National Institute of Informatics –
Tokyo, JP
- Bart Jansen
TU Eindhoven, NL
- Mark R. Jerrum
Queen Mary University of
London, GB
- Eun Jung Kim
University Paris-Dauphine, FR
- Christian Komusiewicz
Universität Marburg, DE
- Martin Koutecký
Technion – Haifa, IL
- Stefan Kratsch
HU Berlin, DE
- Ariel Kulik
Technion – Haifa, IL
- Euiwoong Lee
New York University, US
- Bingkai Lin
National Institute of Informatics –
Tokyo, JP
- Andrea Lincoln
MIT – Cambridge, US
- Daniel Lokshtanov
University of California –
Santa Barbara, US
- Pasin Manurangsi
University of California –
Berkeley, US
- Dániel Marx
Hungarian Academy of Sciences –
Budapest, HU
- Matthias Mnich
Universität Bonn, DE
- Jesper Nederlof
TU Eindhoven, NL
- Rolf Niedermeier
TU Berlin, DE
- Shmuel Onn
Technion – Haifa, IL
- Fahad Panolan
University of Bergen, NO
- Marcin Pilipczuk
University of Warsaw, PL
- Michal Pilipczuk
University of Warsaw, PL
- M. S. Ramanujan
University of Warwick –
Coventry, GB
- Peter Rossmanith
RWTH Aachen, DE
- Ignasi Sau Valls
CNRS – Montpellier, FR
- Saket Saurabh
Institute of Mathematical
Sciences – Chennai, IN
- Hadas Shachnai
Technion – Haifa, IL
- Roohani Sharma
Institute of Mathematical
Sciences – Chennai, IN
- Stefan Szeider
TU Wien, AT
- Dimitrios M. Thilikos
CNRS – Montpellier, FR
- Magnus Wahlström
Royal Holloway, University of
London, GB
- Robert Weismantel
ETH Zürich, CH
- Meirav Zehavi
Ben Gurion University –
Beer Sheva, IL



Report from Dagstuhl Seminar 19042

Practical Yet Composably Secure Cryptographic Protocols

Edited by

Jan Camenisch¹, Ralf Küsters², Anna Lysyanskaya³, and
Alessandra Scafuro⁴

- 1 Dfinity Foundation – Zug, CH, jan@dfinity.org
- 2 Universität Stuttgart, DE, ralf.kuesters@sec.uni-stuttgart.de
- 3 Brown University – Providence, US, anna@cs.brown.edu
- 4 North Carolina State University – Raleigh, US, ascafur@ncsu.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19042 “Practical Yet Composably Secure Cryptographic Protocols”.

The workshop’s main aim was to enhance the community’s understanding of (1) what a good model was for how various protocols and systems co-exist in a larger system; (2) how to model important tasks and security protocols in such a model; (3) how to prove security of protocols in such a model.

Seminar January 20–25, 2019 – <http://www.dagstuhl.de/19042>

2012 ACM Subject Classification Security and privacy → Cryptography

Keywords and phrases applied cryptography, cryptographic protocols, practical protocols, provably secure protocols, security models, universally composability

Digital Object Identifier 10.4230/DagRep.9.1.88

Edited in cooperation with Sophia Yakubov

1 Executive Summary

Jan Camenisch

Ralf Küsters

Anna Lysyanskaya

Alessandra Scafuro

License  Creative Commons BY 3.0 Unported license

© Jan Camenisch, Ralf Küsters, Anna Lysyanskaya, and Alessandra Scafuro

We began by having survey talks on four research threads that had laid foundations of such models. Specifically, Ran Canetti presented his Universal Composability model, Dennis Hofheinz presented his work on the GNUC model, Ralf Küsters presented his IITM/iUC model, and Ueli Maurer presented the model of Constructive Cryptography.

Following these tutorials, we had several talks on how specific security goals and protocols are modeled and proved secure. Björn Tackmann presented a way to model a zero-knowledge proof protocol that made statements about knowledge of certain inputs to ideal functionalities. Manu Drijvers presented a way to model the global random oracle that can be used by participants in different protocols in a composable way.

Once the details of the specific models and how to use them were fresh in everyone’s minds, we split up into working groups. In order to do this, we first had a discussion on what problems we believed were worth tackling; we proposed many problems, and then agreed to discuss a subset of them.



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

Practical Yet Composably Secure Cryptographic Protocols, *Dagstuhl Reports*, Vol. 9, Issue 1, pp. 88–103

Editors: Jan Camenisch, Ralf Küsters, Anna Lysyanskaya, and Alessandra Scafuro



Dagstuhl Reports

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

The topics explored by the working groups are discussed in detail below, in the “results” section of this report. The following additional topics were proposed for discussion (but were not discussed):

- Model asynchrony and time
- Anonymous communication
- Global random oracles in CC
- Secure Message Transfer in various model
- Concrete security in UC/IITM
- Finalise \mathcal{F}_{sig} (with reasons why certain choices are better than others)

Additionally, we had several talks on recent and ongoing research projects. Marc Fischlin on composition of key agreement; Markulf Kohlweiss on structuring game-based proofs; Ran Cohen on probabilistic termination in cryptographic protocols; Antigoni Polychandrou presented two-round two-party computation; Vassilis Zikas modeling the public ledger functionality; Ran Canetti talking about using the EasyCrypt software to aid in cryptographic proofs and verification.

The following is a summary of the workshop results:

1. The relationship between the UC and IITM model was intensively discussed, concluding that the models are very close and that it is possible to unify the two models. The unification also seamlessly includes JUC, GUC, and SUC.
2. The working group on SNARKs (recursive composition of succinct proofs) achieved initial modeling success and crystallization of what’s actually challenging.
3. The working group on modeling \mathcal{F}_{vrf} and constricting it from $\mathcal{F}_{\text{sig}}, \mathcal{F}_{\text{ro}}$ figured out what the stumbling blocks were and what was fundamental.
4. The working group on $\mathcal{F}_{\text{NIZK}}$ and proofs about signatures in Constructive Crypto started to model typical UC functionality in the Constructive Crypto framework and then inspected how they could be composed.
5. The working group on building threshold primitives from single primitive (e.g. threshold signatures from signatures, threshold encryption from encryption etc) came up with a candidate for a “thresholdizer” functionality, and found some subtleties in defining threshold behavior in the ideal world. They also found a candidate construction to test the validity of the definition.
6. The working group on setup assumptions analyzed the assumptions used for constructing composable protocols in terms of practicality and security provided.
7. The working group on delegating secret keys – discovered a simple interface that can be added to \mathcal{F}_{sig} to make it possible to delegate from one user to another well-defined user. Next steps are to investigate if it generalizes to other functionalities and to delegation that’s based on knowledge transfer rather than explicit authorization of identity.

2 Table of Contents

Executive Summary

Jan Camenisch, Ralf Küsters, Anna Lysyanskaya, and Alessandra Scafuro 88

Overview of Talks

Universally Composable Security: Philosophy, History, Status (Or: Exorcising the devil of detail: A never-ending task)

Ran Canetti 91

Probabilistic Termination and Composability of Cryptographic Protocols

Ran Cohen 91

The Wonderful World of Global Random Oracles

Manu Drijvers 92

Game-based Composition for Key Exchange

Marc Fischlin 93

A Bite of GNUC

Dennis Hofheinz 93

State-Separating Proofs for Code-Based Game-Playing Proofs

Markulf Kohlweiss 93

Constructive cryptography and discrete system theory

Ueli Maurer 94

The IITM Model and its Instantiation iUC: Simple and Expressive Universal Composability

Ralf Küsters 95

Multi-protocol UC and its Use for Building Modular and Efficient Protocols

Björn Tackmann 95

Working groups

Extending the UC Signature Functionality with Unpredictability and Applications to Verifiable Random Functions

Markulf Kohlweiss 96

Relating the UC and IITM Models

Ralf Küsters, Ran Canetti, Celine Chevalier, Daniel Rausch, and Björn Tackmann 97

Extending \mathcal{F}_{sig} to Allow for Key Delegation

Anna Lysyanskaya, Celine Chevalier, and Sophia Yakoubov 98

Modeling in the Constructive Cryptography Model

Ueli Maurer, Jan Camenisch, Celine Chevalier, Jens Groth, and Daniel Rausch . . 99

Modular Realization of Threshold Primitives

*Alessandra Scafuro, Stephan Krenn, Ralf Küsters, Daniel Slamanig, and Ivan Visconti*100

Setup Assumptions for Universal Composability

Alessandra Scafuro, Manu Drijvers, Stephan Krenn, Arpita Patra, Antigoni Polychroniadou, and Daniel Slamanig 100

Participants 103

3 Overview of Talks

3.1 Universally Composable Security: Philosophy, History, Status (Or: Exorcising the devil of detail: A never-ending task)

Ran Canetti (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Ran Canetti

Main reference Ran Canetti: “Universally Composable Security: A New Paradigm for Cryptographic Protocols”, in Proc. of the 42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA, pp. 136–145, IEEE Computer Society, 2001.

URL <https://doi.org/10.1109/SFCS.2001.959888>

This talk contained an overview of the notion of Universally Composable (UC) security, its history, variants, and current status.

3.2 Probabilistic Termination and Composability of Cryptographic Protocols

Ran Cohen (MIT – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Ran Cohen

Joint work of Sandro Coretti, Juan Garay and Vassilis Zikas

Main reference Ran Cohen, Sandro Coretti, Juan A. Garay, Vassilis Zikas: “Probabilistic Termination and Composability of Cryptographic Protocols”, in Proc. of the Advances in Cryptology – CRYPTO 2016 – 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III, Lecture Notes in Computer Science, Vol. 9816, pp. 240–269, Springer, 2016.

URL https://doi.org/10.1007/978-3-662-53015-3_9

Main reference Ran Cohen, Sandro Coretti, Juan A. Garay, Vassilis Zikas: “Round-Preserving Parallel Composition of Probabilistic-Termination Cryptographic Protocols”, in Proc. of the 44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland, LIPIcs, Vol. 80, pp. 37:1–37:15, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2017.

URL <http://dx.doi.org/10.4230/LIPIcs.ICALP.2017.37>

Since the introduction of secure multiparty computation (MPC) in the ’80s, it has been a common practice to consider a broadcast channel when designing MPC protocols. Well-known lower bounds show that deterministic broadcast protocols cannot run in a number of rounds sub-linear in the number of corrupted parties. The seminal works of Ben-Or and Rabin showed how to overcome these limitations via randomization, igniting the study of protocols over point-to-point channels with *probabilistic termination (PT)* and expected constant round complexity. However, absent a rigorous simulation-based definition, the suggested protocols are proven secure in a property-based manner, and therefore guarantee limited, if any, composability.

Composing PT protocols affects the round complexity of the resulting protocol in somewhat unexpected ways. For instance, the expected round complexity of the parallel composition of expected-constant-round protocols might be logarithmic in number of instances. Sequential composition of PT protocol also raises subtle issues since the parties fall out-of-sync and cannot start the protocol at the same round.

In this work, we put forth the first simulation-based treatment of MPC with probabilistic termination in the UC framework and prove a universal composition theorem for PT protocols. Our theorem allows one to compile a protocol using deterministic-termination hybrids into a protocol that uses expected-constant-round protocols for emulating these hybrids, preserving

the expected round complexity of the calling protocol. We showcase our definitions and compiler by providing the first composable protocols (with simulation-based security proofs) over point-to-point channels for the following primitives: (1) expected-constant-round perfect Byzantine agreement, (2) expected-constant-round perfect parallel broadcast, and (3) MPC with round complexity independent of the number of parties.

We proceed to analyze whether the techniques used for parallel composition of broadcast (which is a privacy-free functionality) can be generalized for composing in parallel arbitrary MPC protocols, and provide both feasibility and infeasibility results. We show an efficient protocol-compiler that outputs a protocol that realizes the parallel composition of m protocols, without increasing the expected round complexity; moreover, the compiler requires only black-box access to the underlying *protocols*. Using known techniques, a similar result cannot be achieved given only black-box access to the *functions* realized by the protocols.

3.3 The Wonderful World of Global Random Oracles

Manu Drijvers (Dfinity – Zürich, CH)

License  Creative Commons BY 3.0 Unported license

© Manu Drijvers

Joint work of Manu Drijvers, Jan Camenisch, Tommaso Gagliardoni, Anja Lehmann, Gregory Neven
Main reference Jan Camenisch, Manu Drijvers, Tommaso Gagliardoni, Anja Lehmann, Gregory Neven: “The Wonderful World of Global Random Oracles”, in Proc. of the Advances in Cryptology – EUROCRYPT 2018 – 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 – May 3, 2018 Proceedings, Part I, Lecture Notes in Computer Science, Vol. 10820, pp. 280–312, Springer, 2018.

URL http://dx.doi.org/10.1007/978-3-319-78381-9_11

The random-oracle model by Bellare and Rogaway (CCS’93) is an indispensable tool for the security analysis of practical cryptographic protocols. However, the traditional random-oracle model fails to guarantee security when a protocol is composed with arbitrary protocols that use the same random oracle. Canetti, Jain, and Scafuro (CCS’14) put forth a global but non-programmable random oracle in the Generalized UC framework and showed that some basic cryptographic primitives with composable security can be efficiently realized in their model. Because their random-oracle functionality is non-programmable, there are many practical protocols that have no hope of being proved secure using it. In this paper, we study alternative definitions of a global random oracle and, perhaps surprisingly, show that these allow one to prove GUC-secure existing, very practical realizations of a number of essential cryptographic primitives including public-key encryption, non-committing encryption, commitments, Schnorr signatures, and hash-and-invert signatures. Some of our results hold generically for any suitable scheme proven secure in the traditional ROM, some hold for specific constructions only. Our results include many highly practical protocols, for example, the folklore commitment scheme $H(m|r)$ where m is a message and r is the random opening information.

3.4 Game-based Composition for Key Exchange

Marc Fischlin (TU Darmstadt, DE)

License © Creative Commons BY 3.0 Unported license
© Marc Fischlin

Joint work of Marc Fischlin, Chris Brzuska, Bogdan Warinschi, Felix Günther

We discuss composition of key exchange protocols with arbitrary symmetric-key protocols, like a secure channel. We use game-based security notions for the primitives and the composed protocol. It turns out that the secure composition requires specific properties of the key exchange protocol such as forward security and public matching of partnered sessions.

3.5 A Bite of GNUC

Dennis Hofheinz (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license
© Dennis Hofheinz

Joint work of Dennis Hofheinz, Victor Shoup

Main reference Dennis Hofheinz, Victor Shoup: “GNUC: A New Universal Composability Framework”, J. Cryptology, Vol. 28(3), pp. 423–508, 2015.

URL <https://doi.org/10.1007/s00145-013-9160-y>

GNUC (for “GNUC’s not UC”) is a framework that allows to model multi-party protocols and to analyze their security properties. This talk highlights two technical design choices made in the GNUC universal composability framework:

- the notion of “efficiency” for protocol machines, and
- the organization of protocols machines in a hierarchical manner.

In particular, we explain that the notion of an “efficient” protocol (or of an “efficient” attack or distinguisher) in fact contains a number of technical pitfalls, and how we avoid those pitfalls in GNUC.

The take-away message of this talk should be

- that there are lots of low-level decisions to be taken when designing a framework for multi-party protocols,
- that these low-level decisions may have high-level consequences (e.g., for the expressiveness or security properties of the resulting framework),
- but that these low-level decisions should not distract from the high-level proof ideas one usually tries to convey when using protocol frameworks.

3.6 State-Separating Proofs for Code-Based Game-Playing Proofs

Markulf Kohlweiss (University of Edinburgh, GB)

License © Creative Commons BY 3.0 Unported license
© Markulf Kohlweiss

The security analysis of real-world protocols involves reduction steps that are conceptually simple but still have to account for many protocol complications found in standards and implementations. Taking inspiration from universal composability, abstract cryptography, process algebras, and type-based verification frameworks, we propose a method to simplify

large reductions, avoid mistakes in carrying them out, and obtain concise security statements. Our method decomposes monolithic games into collections of stateful packages representing collections of oracles that call one another using well-defined interfaces. Every component scheme yields a pair of a real and an ideal package. In security proofs, we then successively replace each real package with its ideal counterpart, treating the other packages as the reduction. We build this reduction by applying a number of algebraic operations on packages justified by their state separation. Our method handles reductions that emulate the game perfectly, and leaves more complex arguments to existing game-based proof techniques such as the code-based analysis suggested by Bellare and Rogaway. It also facilitates computer-aided proofs, inasmuch as the perfect reductions steps can be automatically discharged by proof assistants. We illustrate our method on two generic composition proofs: (1) a proof of self-composition using a hybrid argument; and (2) the composition of keying and keyed components. For concreteness, we apply them to the KEM-DEM proof of hybrid-encryption by Cramer and Shoup and to the composition of forward-secure game-based key exchange protocols with symmetric-key protocols.

3.7 Constructive cryptography and discrete system theory

Ueli Maurer (ETH Zürich, CH)

License  Creative Commons BY 3.0 Unported license
© Ueli Maurer

Joint work of Ueli Maurer, Renato Renner

The talk presented three parts.

1. An abstract resource theory, where resources are elements of a partially ordered set and constructions (of resources from resources), adjoining resources, relaxations, and several other concepts are captured by order-preserving functions (i.e., homomorphisms) satisfying certain axioms, for example (one-sided) commutativity of certain homomorphisms.
2. A theory of discrete probabilistic systems, where most systems discussed in cryptography can be understood as descriptions (in a particular language specific to the context and paper, for example a specific pseudo-code language) of such discrete systems. One can consider system specifications (i.e., sets of probabilistic systems) and define various specification relaxations, including an ϵ -relaxation and the game-relaxation of a system containing a game, where the relaxation is defined as the set of systems behaving like the given system but where nothing is specified if the game is won.
3. Constructive cryptography as an instantiation of a resource theory instantiated with discrete systems. Many examples were presented. A specific example that can probably not be captured by previous concepts in cryptography is authentication amplification, meaning that one constructs an n -bit authenticated channel from a k -bit authenticated channel (for $n > k$) by use of a hash function. By explaining this using game-relaxation one can make a tight construction statement without need for reductions or a distinguisher or adversary concept, and despite the fact that a single hash function is never collision-resistant.

Joint work with Renato Renner, and also based on joint work and discussions with many other people, in particular Björn Tackmann.

3.8 The IITM Model and its Instantiation iUC: Simple and Expressive Universal Composability

Ralf Küsters (Universität Stuttgart, DE)

License © Creative Commons BY 3.0 Unported license
© Ralf Küsters

Joint work of Ralf Küsters, Max Tuengerthal, Daniel Rausch, Jan Camenisch, Stephan Krenn

Main reference Ralf Küsters, Max Tuengerthal: “The IITM Model: a Simple and Expressive Model for Universal Composability”, IACR Cryptology ePrint Archive, Vol. 2013, p. 25, 2013.

URL <http://eprint.iacr.org/2013/025>

The universal composability paradigm allows for the modular design and analysis of cryptographic protocols. It has been widely and successfully used in cryptography. However, devising a coherent yet simple and expressive model for universal composability is, as the history of such models shows, highly non-trivial.

In this tutorial, we present a coherent model for universal composability, called the IITM model (“Inexhaustible Interactive Turing Machine”). A main feature of the model is that it is stated without a priori fixing irrelevant details, such as a specific way of addressing of machines by session and party identifiers, a specific modeling of corruption, or a specific protocol hierarchy. In addition, we employ a very general notion of runtime. All reasonable protocols and ideal functionalities should be expressible based on this notion in a direct and natural way, and without tweaks, such as (artificial) padding of messages or (artificially) adding extra messages.

The expressivity of the IITM is also reflected in the fact that joint-state and global state composition theorems follow directly from the basic composition theorem of the IITM model. No model extensions or new theorems are necessary. The model also allows for modeling forms of shared state that are out of reach of other models. Moreover, protocols can be modeled where protocol participants are not forced to establish session IDs before the start of the protocol.

Finally, we briefly discuss an instantiation of the IITM model, called iUC, which helps protocol designers in their modeling and analysis tasks.

IITM: <http://eprint.iacr.org/2013/025/> (joint work with Max Tuengerthal and Daniel Rausch) iUC: Will soon be made available on eprint (joint work with Jan Camenisch, Stephan Krenn, and Daniel Rausch)

3.9 Multi-protocol UC and its Use for Building Modular and Efficient Protocols

Björn Tackmann (IBM Research-Zurich, CH)

License © Creative Commons BY 3.0 Unported license
© Björn Tackmann

Main reference Jan Camenisch, Manu Drijvers, Björn Tackmann: “Multi-Protocol UC and its Use for Building Modular and Efficient Protocols”, IACR Cryptology ePrint Archive, Vol. 2019, p. 65, 2019.

URL <https://eprint.iacr.org/2019/065>

We want to design and analyze protocols in a modular way by combining idealized components that we realize individually. While this is in principle possible using security frameworks that provide generic composition theorems, we notice that actually applying this methodology in practical protocols is far from trivial and, worse, is sometimes not even possible. As an example, we use a natural combination of zero-knowledge proofs with signature and

commitment schemes, where the goal is to have a party prove in zero-knowledge that it knows a signature on a committed message, i.e., prove knowledge of a witness to a statement involving algorithms of the signature and commitment scheme. We notice that, unfortunately, the composition theorem of the widely used UC framework does not allow one to modularly prove the security of this example protocol.

We then describe a new variant of the UC framework, multi-protocol UC, and show a composition theorem that generalizes the one from the standard framework. We use this new framework to provide a modular analysis of a practical protocol that follows the above structure and is based on discrete-logarithm-based primitives. Besides the individual security proofs of the protocol components, we also describe a new methodology for idealizing them as components that can then be composed.

4 Working groups

4.1 Extending the UC Signature Functionality with Unpredictability and Applications to Verifiable Random Functions

Markulf Kohlweiss (University of Edinburgh, GB)

License  Creative Commons BY 3.0 Unported license
© Markulf Kohlweiss

Joint work of Markulf Kohlweiss, Dennis Hofheinz, Anna Lysyanskaya, Marc Fischlin, Manu Drijvers, Vassilis Zikas, Celine Chevalier

The universal composability (UC) framework guarantees that a protocol remains secure even when composed with arbitrary other protocols. A composition theorem allows UC secure protocols to be built iteratively by composing protocols that already have been proven UC secure. The proof is performed in a hybrid world with ideal functionalities that describe the guarantees of the component protocols. One such ideal functionality is the \mathcal{F}_{Sig} functionality that models the security of existentially unforgeable signatures (EUF-CMA).

EUF-CMA secure signatures are by necessity unpredictable. That is, no efficient adversary can compute a valid signature except with negligible probability. This property, however, is not modeled by any existing formulation of the \mathcal{F}_{Sig} functionality. Existing formulations typically ask the simulator to determine how signatures are computed, either by asking for the signature itself or for signing algorithms that generate the signature.

This state of affairs limits the usefulness of \mathcal{F}_{Sig} in settings where the unpredictability of signatures is essential, e.g. for the construction of verifiable random functions. A verifiable random function scheme fixes a family of functions $\{f_{\text{sk}}\}$ and a way to sample a public key pair (sk, pk) . The secret key sk is used to evaluate the function on a point x to obtain $y = f_{\text{sk}}(x)$ and a proof π . The public key is used in a verification algorithm $\text{Verify}(\text{pk}, x, y, \pi)$. A secure VRF satisfies two properties:

1. without seeing π , y is indistinguishable from random for any efficient adversary,
2. an efficient adversary cannot compute x, y, π such that $\text{Verify}(\text{pk}, x, y, \pi) = 1$.

Ideally given a functionality for unpredictable signatures \mathcal{F}_{Sig} we would like to be able to construct a UC protocol emulating a UC functionality for \mathcal{F}_{Vrf} in the random oracle model. This problem is interesting for several reasons:

1. VRFs are an important building block of proof-of-stake ledger (aka blockchain) protocols
2. It points out a current weakness of proof techniques in the UC model with regard to reasoning about rare events in UC, such as reasoning about signature forgeries, hash function collisions, or dishonest majorities.

A technique we explored was to adapt functionalities that commit “suicide” by making themselves trivially distinguishable from the real world into functionalities that become “super-useful”. This would make the functionality useful when used in a hybrid proof, in which the suicide in case of the rare event would make it useless otherwise. These techniques appear promising, but did not suffice to construct VRF from unpredictable signatures. Positive (side) outcomes of the discussion were

- a better understanding of a recent \mathcal{F}_{vrf} functionality. The published variant was buggy and we managed to contact the authors to track down the issue. The paper was subsequently updated.
- identification of interesting research questions: 1. constructing \mathcal{F}_{vrf} with strong properties in the standard model, 2. idealization of computational entropy using actual entropy.
- an alternative technique for constructing \mathcal{F}_{vrf} using signatures that can be split into an entropy part and a proof part.

4.2 Relating the UC and IITM Models

Ralf Küsters (Universität Stuttgart, DE), Ran Canetti (Tel Aviv University, IL), Celine Chevalier (University Paris II, FR), Daniel Rausch (Universität Stuttgart, DE), and Björn Tackmann (IBM Research-Zurich, CH)

License © Creative Commons BY 3.0 Unported license
© Ralf Küsters, Ran Canetti, Celine Chevalier, Daniel Rausch, and Björn Tackmann

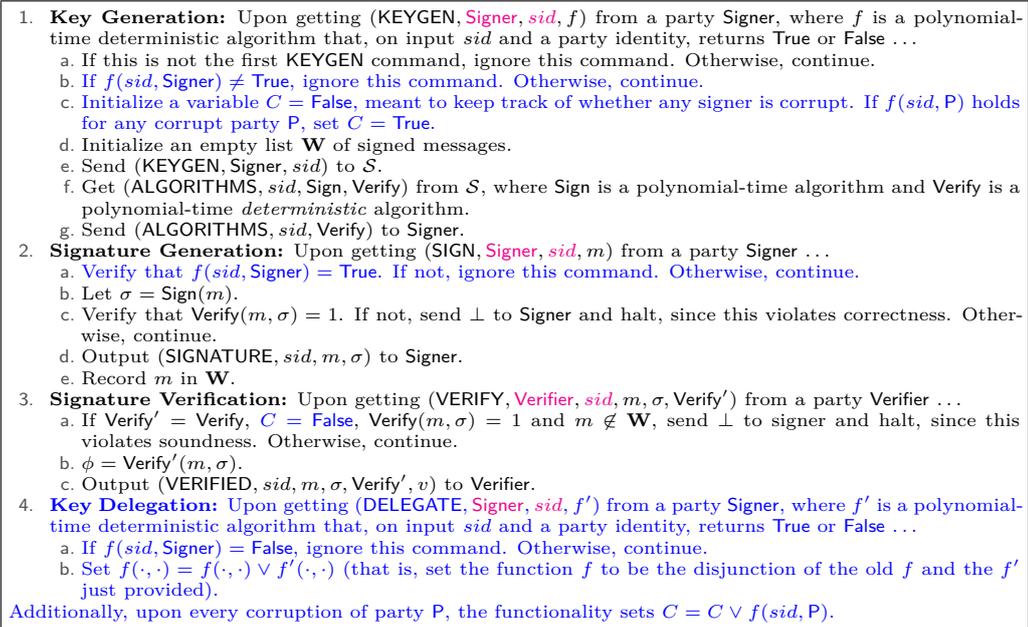
Two of the most prominent models for analyzing protocols in a universally composable manner are the UC model by Canetti [1] and the IITM model by Kuesters et al. [2] At a very high level, both models follow a similar idea and provide theorems for the secure composition of protocols. However, at a technical level, both models are (sometimes drastically) different. This includes the machine model, connections between machines, addressing of machines, runtime definitions, as well as requirements of the composition theorems. The goal of this working group was to relate both models on this technical level and find a common ground.

We started our discussion by trying to find a common set of rules for protocols which, if they are met, imply that the protocol can be analyzed, proven secure, and composed in both the UC and the IITM model. This meant we had to find the limits of each model, and see whether there are certain types of protocols or features that can be expressed only in one model. Surprisingly, during this discussion, we found out that both models are actually closer related than expected. For (almost) every technical aspect and way of modeling a protocol in one model, we found a way to achieve the same in the other model. We gained many interesting insights in how the same problems are solved in different yet equivalent ways by each model.

We decided that we want to collaborate on a paper as a followup to this working group. The paper shall explore the insights from this working group in more detail and show that the UC model and the IITM model are actually equivalent in terms of expressivity (up to runtime). This has many interesting consequences. For example, it would follow that, as already shown in the IITM model, the composition theorem in the UC model also implies theorems for joint-state and global state composition as special cases.

References

- 1 Ran Canetti. *Universally Composable Security: A New Paradigm for Cryptographic Protocols*. FOCS, 2001. See <https://eprint.iacr.org/2000/067> for the most recent version.
- 2 Ralf Küsters. *Simulation-Based Security with Inexhaustible Interactive Turing Machines*. CSFW, 2006. See <https://eprint.iacr.org/2013/025> for the most recent version.



■ **Figure 1** Ideal Functionality for Digital Signatures With Key Delegation.

4.3 Extending \mathcal{F}_{sig} to Allow for Key Delegation

Anna Lysyanskaya (Brown University – Providence, US), Celine Chevalier (University Paris II, FR), and Sophia Yakoubov (MIT Lincoln Laboratory – Lexington, US)

License  Creative Commons BY 3.0 Unported license
 © Anna Lysyanskaya, Celine Chevalier, and Sophia Yakoubov

In Figure 1, we describe the signature functionality \mathcal{F}_{sig} , with the modification that a signer can delegate their signing ability (e.g. by sharing their secret key) to others. This delegation module can be similarly grafted onto other functionalities, e.g. encryption.

Note that currently, this functionality only describes identity-based delegation (that is, a signer must specify a function on users' identities which determines whether they have the power to sign). Ultimately, it would be desirable to make this function more generic, enabling witness-based delegation.

In **magenta**, we denote inputs to the ideal functionality that a party (controlled by the environment / adversary) cannot falsify. In **blue**, we denote places where this functionality differs from the digital signature ideal functionality of Ran Canetti's Universal Composability paper (2005 version).

Future work will include:

1. Grafting this module onto other functionalities.
2. Exploring witness-based delegation.

4.4 Modeling in the Constructive Cryptography Model

Ueli Maurer (ETH Zürich, CH), Jan Camenisch (Dfinity Foundation – Zug, CH), Celine Chevalier (University Paris II, FR), Jens Groth (London, GB), and Daniel Rausch (Universität Stuttgart, DE)

License  Creative Commons BY 3.0 Unported license
© Ueli Maurer, Jan Camenisch, Celine Chevalier, Jens Groth, and Daniel Rausch

The Constructive Cryptography (CC) model by Maurer [1] is a framework for performing modular proofs. Compared to the widely used Universal Composability (UC) model [2], the CC model takes a much more abstract view and defines only a minimal number of details that are sufficient to obtain modularity. In particular, the CC model does not specify a specific computational model or a runtime notion. Thus it is not directly obvious how ideal functionalities and security proofs as used in the UC model can be modeled in the CC model.

Our working group had two goals. Firstly, we wanted to show that (and how) one can model functionalities and perform security proofs from the UC model also in the CC model, thus verifying the expressiveness of the CC model and making the CC model more accessible to cryptographers that are used to the UC model. Secondly, we wanted to find ways to use the more abstract view of the CC framework for simplifying or generalizing functionalities as well as making security proofs easier to carry out and verify. We used a recent protocol by Jan Camenisch et al. [3] defined in the UC model as a case study. This protocol combines ideal functionalities for zero knowledge (ZK), commitments, and digital signatures in a modular fashion to prove knowledge of a signature on a message in a commitment.

We started by defining an ideal functionality for digital signatures in the CC model. A core insight of this process was that the realization proof of the signature functionality becomes much simpler in the CC model. This is because one can postpone the final reduction to a game based security assumption until after the whole system has been designed, whereas this reduction must be performed as part of the realization proof in the UC model. In the next step, we defined an ideal functionality for non-interactive ZK proofs of knowledge in the CC model and sketched a realization, including a corruption model. While this first version of the ZK functionality does not appear to be much simpler than in the UC model, it illustrates how cryptographers can express complex protocols from the UC model in the more abstract CC model.

The results of our working group indicate that we can indeed bridge the gap between the UC and CC model. Furthermore, our experience with the signature functionality makes us confident that we can find additional improvements to simplicity and generality of ideal functionalities and security proofs. Thus, our team decided to follow up in this working group with a paper based on the insights that we gained.

References

- 1 Ueli Maurer. *Constructive Cryptography – A New Paradigm for Security Definitions and Proofs*. Theory of Security and Applications – Joint Workshop, TOSCA, 2011.
- 2 Ran Canetti. *Universally Composable Security: A New Paradigm for Cryptographic Protocols*. FOCS, 2001. See <https://eprint.iacr.org/2000/067> for the most recent version.
- 3 Jan Camenisch, Manu Drijvers, and Björn Tackmann. *Multi-Protocol UC and its Use for Building Modular and Efficient Protocols*. IACR Cryptology ePrint Archive, 2019. Available at <https://eprint.iacr.org/2019/065>.

4.5 Modular Realization of Threshold Primitives

Alessandra Scafuro (North Carolina State University – Raleigh, US), Stephan Krenn (AIT – Austrian Institute of Technology – Wien, AT), Ralf Küsters (Universität Stuttgart, DE), Daniel Slamanig (AIT – Austrian Institute of Technology – Wien, AT), and Ivan Visconti (University of Salerno, IT)

License © Creative Commons BY 3.0 Unported license
© Alessandra Scafuro, Stephan Krenn, Ralf Küsters, Daniel Slamanig, and Ivan Visconti

Research Question. In a (t, n) threshold primitive we want that a cryptographic operation is performed if a subset of t parties agree (and has the credential) to perform a certain operation on a input x . For example, threshold signature require that at least t parties agree on signing a message, threshold encryption requires that at least t parties participate to decrypt a ciphertext c .

The motivation behind threshold schemes is typically robustness, i.e., we want to make sure that if one of more machines fail, the security of a certain operation is still guaranteed. In some settings (such as ring signatures) we additionally want privacy, and we require that the identity of the identity of the t participants is not leaked. We are not considering this setting here.

Definitions of threshold encryption and (ring) signatures exist, in a game-based setting, and only very recently a definition of threshold (ring) signature was also provided as ideal functionality.

Our goal is generalize the problem of computing threshold X having access to primitive X . We want to provide a “wrapper” ideal functionality \mathcal{F}_{TR} that captures the threshold constraint wrt an operation, making calls to the underlying functionality \mathcal{F}_{OP} that performs a single operation.

Progress made over the meeting. We have outlined an ideal threshold signature functionality that has access to an instance of signature functionality. During this process we identified potential issues in current definitions of threshold ring signatures and identified edge cases that do not seem to be explicitly captured by current definitions. We also discussed a candidate, proof of concept realization of our new ideal functionality, based on signatures, Merkle tree and succinct non-interactive zero-knowledge argument.

Next step. Our next step is to formally validate our high-level designs.

4.6 Setup Assumptions for Universal Composability

Alessandra Scafuro (North Carolina State University – Raleigh, US), Manu Drijvers (Dfinity – Zürich, CH), Stephan Krenn (AIT – Austrian Institute of Technology – Wien, AT), Arpita Patra (Indian Institute of Science – Bangalore, IN), Antigoni Polychroniadou (Cornell Tech – New York, US), and Daniel Slamanig (AIT – Austrian Institute of Technology – Wien, AT)

License © Creative Commons BY 3.0 Unported license
© Alessandra Scafuro, Manu Drijvers, Stephan Krenn, Arpita Patra, Antigoni Polychroniadou, and Daniel Slamanig

Research Question. Setup assumptions are assumptions we make about the world. One example of setup is the assumption that the world has trustworthy parties that can honestly perform certain operations, or that users are able to manufacture physical tokens that behave like black-boxes and completely obfuscate a software.

Setup assumptions are necessary for proving security of protocol in universally composable (UC) sense, thus they are extensively used in any UC-secure protocol.

Naturally, some assumptions might be more practical/realistic than others, at the expenses of requiring more trust on third parties and hence more assumptions about the (honest) behaviour we expect in the real world.

In particular, assuming setups that have local scopes (i.e., they are visible only to the parties participating in the protocol) is quite unrealistic, as in the real world one would assume that the same setup (e.g., the same public key) is re-used in many executions. On the other hands, using global setups, while seemingly a more realistic approach, it introduces global trapdoors that could determine global failure in case the trusted party is compromised.

The literature seem to lack of a thorough analysis and comparison of the existent setup assumptions, especially, in light of global composition requirements.

Progress made over the meeting. During the meeting we discussed a few setup assumptions used in the literature and compared them wrt two parameters: practicality, that is, how easily they could be realized, and trust.

- **Common Reference String (CRS model).** This setup assumes that for each protocol execution, the participating parties are able to obtain a local common reference string (sampled from a distribution prescribed by the protocol) from a trusted source. The string is fresh and local to the protocol. In the literature, it is typically assumed that the common reference string is sampled by a trusted party. Alternatively, the string can be computed by a set of parties, via a multiparty computation protocol. Variation of the CRS model exist, however such models only relax the requirement of trusting a single source. Note that the requirement of the CRS being local and fresh is necessary only when proving UC-security. If weaker composability guarantees are required (e.g, simply proving that a protocol is a non-interactive zero-knowledge protocol) then the same CRS could be reused and have a global scope.
 - **Practicality:** mostly impractical. Since the CRS must be local and freshly sampled upon each protocol execution (or a predetermined set of executions), parties will need continue access to a trusted source of CRS. If the CRS were computed via a MPC protocol, this process would be even more cumbersome,
 - **Trust.** If maliciously computed the CRS contains trapdoors. Corruption of a CRS source determines loss of security. However, since each protocol is executed with fresh CRS, corruption of a one CRS source has only a local impact and it affects only one execution (assuming that different protocols might have access to different CRS sources).
- **Augmented Common Reference String.** In this setup, there is a trusted party that publishes a short reference string (i.e., of size independent on the number of parties), and possesses a master secret key that can be used to derive per-party secret keys. Only the reference string is required for executing a protocol. However, corrupted parties can ask the trusted party to provide them with their secret key. In the security proof this secret key is used by the simulator.
- **Key Registration Model.** In this setup, there is a trusted party that chooses a private and public key pair for each registered party and lets all parties know the value of the public key only. Corrupted parties however can retrieve their own secret key by asking the trusted party.

- Practicality. Somewhat practical. The key registration authority could be implemented by a third party (post office is often suggested). A potential practical drawback is that the KR authority must be always on-line for registration and key retrieval. This means that a MPC emulation of the KR authority might be problematic since all parties must be on-line. On the other hand, perhaps an MPC implementation would discourage people to cheat since so many parties will then learn that this person is asking for their key.
- Trust. The key registration authority knows all trapdoors of the system that are used in all executions. As such it represents a single point of global failure.
- Similar arguments hold for the Augmented Common Reference String model.
- Global Random Oracle. This is a proof methodology rather than a setup assumption since there is no real-world implementation of the random oracle. The global random oracle assumption idealizes properties of hash functions. It assumes that all parties have oracle access to a public random function. In the proof this is translated with the ability of the simulator to extract the queries made to the oracle. There exist two formulations of the Random Oracle Model, one assumes that random oracles are local to the protocol execution. This is somewhat less realistic since in reality the same hash function is used across all execution. The global random oracle model instead assumes that there is a global oracle that all protocol executions share. This limits the power of the simulator in the proof.
 - Practicality. When the RO is implemented with an hash function, it is very practical.
 - Trust/Security. Concrete hash functions do not behave like a random oracle, therefore when RO is replaced with hash function it does not provide provable security guarantees.
- Hardware Assumptions. In this setting parties are assumed to have the ability to manufacture hardware tokens that embed arbitrary functions and behave like a black-box when in the hands of an adversary.
 - Practicality. The exchange and the creation of general purpose hardware token is highly impractical. Some constructions in literature, are based on very specific tokens, such as signature cards, that are easy to obtain from trusted authorities. Such construction are significantly more practical, but their security completely relies on the trust in the authority manufacturing the tokens, somewhat collapsing to the Key Registration Model.
 - Trust. The trust here is posed into the hardware technology as well as the trust into manufacturers.

Concluding thoughts. Known global setups guarantee the highest level of composability, but they seem to provide the most fragile security guarantees since the security of all protocols in the system rely on the security of global trapdoors. If such trapdoors fall into the hands of adversary, the security of every protocol is compromised.

On the contrary, protocols achieving weaker composition guarantees, are less fragile in that they are based on local trapdoors or no trapdoors (e.g., in stand-alone security) This seems to suggest that the highest composability guarantee comes at the price of highest reliance on real world good behaviour of third parties. The global random oracle does not suffer of this problem, since the work that is done by the simulator in the ideal world simply cannot be emulated in the real world by any adversary since random oracles do not exist.

Next steps. The next step would be to write a manuscript analysing more thoroughly possible implementations of each setup assumption, the concrete trust required and the possible “fall-back” security guarantees in presence of compromise.

Participants

- Jan Camenisch
Dfinity Foundation – Zug, CH
- Ran Canetti
Tel Aviv University, IL
- Celine Chevalier
University Paris II, FR
- Ran Cohen
MIT – Cambridge, US
- Manu Drijvers
Dfinity – Zürich, CH
- Marc Fischlin
TU Darmstadt, DE
- Dov Gordon
George Mason University –
Fairfax, US
- Jens Groth
London, GB
- Timo Hanke
Dfinity Foundation – Zug, CH
- Dennis Hofheinz
KIT – Karlsruher Institut für
Technologie, DE
- Markulf Kohlweiss
University of Edinburgh, GB
- Stephan Krenn
AIT – Austrian Institute of
Technology – Wien, AT
- Ralf Küsters
Universität Stuttgart, DE
- Anna Lysyanskaya
Brown University –
Providence, US
- Mary Maller
University College London, GB
- Ueli Maurer
ETH Zürich, CH
- Arpita Patra
Indian Institute of Science –
Bangalore, IN
- Antigoni Polychroniadou
Cornell Tech – New York, US
- Daniel Rausch
Universität Stuttgart, DE
- Alessandra Scafuro
North Carolina State University –
Raleigh, US
- Daniel Slamanig
AIT – Austrian Institute of
Technology – Wien, AT
- Björn Tackmann
IBM Research-Zurich, CH
- Muthuramakrishnan
Venkatasubramaniam
University of Rochester, US
- Ivan Visconti
University of Salerno, IT
- Sophia Yakoubov
MIT Lincoln Laboratory –
Lexington, US
- Vassilis Zikas
University of Edinburgh, GB



Data Structures for the Cloud and External Memory Data

Edited by

Gerth Stølting Brodal¹, Ulrich Carsten Meyer², Markus E. Nebel³,
and Robert Sedgewick⁴

1 Aarhus University, DK, gerth@cs.au.dk

2 Goethe-Universität – Frankfurt a. M., DE, umeyer@cs.uni-frankfurt.de

3 Universität Bielefeld, DE, nebel@techfak.uni-bielefeld.de

4 Princeton University, US, rs@cs.princeton.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16101 “Data Structures for the Cloud and External Memory Data”. In today’s computing environment vast amounts of data are processed, exchanged and analyzed. The manner in which information is stored profoundly influences the efficiency of these operations over the data. In spite of the maturity of the field many data structuring problems are still open, while new ones arise due to technological advances. The seminar covered both recent advances in the “classical” data structuring topics as well as new models of computation adapted to modern architectures, scientific studies that reveal the need for such models, applications where large data sets play a central role, modern computing platforms for very large data, and new data structures for large data in modern architectures. The extended abstracts included in this report contain both recent state of the art advances and lay the foundation for new directions within data structures research.

Seminar January 27 – February 1, 2019 – <http://www.dagstuhl.de/19051>

2012 ACM Subject Classification Theory of computation → Data structures design and analysis,
Theory of computation → Design and analysis of algorithms

Keywords and phrases algorithms, big data, cloud computing, data structures, external memory methods, large data sets, web-scale

Digital Object Identifier 10.4230/DagRep.9.1.104

Edited in cooperation with Manuel Penschuck

1 Executive Summary

Gerth Stølting Brodal (Aarhus University, DK)

Ulrich Carsten Meyer (Goethe-Universität – Frankfurt a. M., DE)

Markus E. Nebel (TU Kaiserslautern, DE)

Robert Sedgewick (Princeton University, US)

License  Creative Commons BY 3.0 Unported license

© Gerth Stølting Brodal, Ulrich Carsten Meyer, Markus E. Nebel, and Robert Sedgewick

About the Seminar

Data structures provide ways of storing and manipulating data and information that are appropriate for the computational model at hand. Every such model relies on assumptions that we have to keep questioning. The aim of this seminar was to exchange ideas for new algorithms and data structures, and to discuss our models of computations in light of recent technological advances. This Dagstuhl seminar was the 13th in a series of loosely related Dagstuhl seminars on data structures.



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

Data Structures for the Cloud and External Memory Data, *Dagstuhl Reports*, Vol. 9, Issue 1, pp. 104–124

Editors: Gerth Stølting Brodal, Ulrich Carsten Meyer, Markus E. Nebel, and Robert Sedgewick



Dagstuhl Reports

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

Topics

The presentations covered both advances in classic fields, as well as new problems and insights for recent trends in computing. In particular, Johnson (Section 3.12) and Muth (Section 3.17) reported on models and research opportunities in the cloud and external memory motivated by practical demands.

A number of talks highlighted technical challenges in storing and processing large datasets: Bast (Section 3.2) demonstrated the knowledge database QLever and discussed algorithmic aspects. Distributed frameworks were presented by Bingmann (Section 3.4) reporting on the progress of Thrill while focusing on parallel external sorting and by Yan (Section 3.32) who introduced G-thinker. Farach-Colton (Section 3.7) analyzed the slow-down of various filesystems caused by updates over time. Owens (Section 3.19) discusses intricacies of GPUs and presented efficient and practical data structures for this hardware.

In order to mitigate the impact of huge datasets, streaming and online algorithms were considered. Martinez (Section 3.15) discussed Affirmative Sampling which takes uniform samples of a stream and adapts the sample size to the stream's diversity. Sedgewick (Section 3.26) revisited the cardinality estimation problem and proposed the HyperBitBit algorithm. A matching of requests to resources in an online setting was covered by Raghvendra (Section 3.22). Similarly, Mehlhorn (Section 3.16) presented a solution to assigning indivisible resources approximately optimizing the social welfare.

Nebel (Section 3.18) and Wild (Section 3.31) proposed and analyzed tree-based data structures. Additionally, various aspects on more general graph processing were covered ranging from their enumeration (Lumbroso, Section 3.14) and random sampling (Penschuck, Section 3.20), over representations for k -connectivity (Pettie, Section 3.21) to the detection of substructures (Silvestri, Section 3.28 and Tarjan, Section 3.29).

Regarding the complexity of graph algorithms, Fagerberg (Section 3.23) presented new lower bounds on the reorganisation cost of B -trees, while Thankachan (Section 3.30) gave hardness results on the recognizability of Wheeler graphs. Kopelowitz (Section 3.13) considered the complexity of data structures for the set-disjointness problem. Emphasizing cloud-related security concerns, Jacob (Section 3.11) showed that a range of simple data structures have to incur an $\Omega(\log n)$ overhead if one wants to prevent information leakage via their access patterns.

Problems involving large text corpora were considered by Fischer (Section 3.8) presenting an external memory bi-directional compression scheme, by Golin (Section 3.9) discussing AIFV codes, and by Salinger (Section 3.24) analyzing persistent full-text indices for versioned documents.

Data structures using hashing were examined by Conway (Section 3.5), Dietzfelbinger (Section 3.6), Even and Sanders (Section 3.25). Bender (Section 3.3) discussed variants of Bloom filters which adapt based on past queries.

Afshani (Section 3.1) presented Fragile Complexity, a novel model of computation with an element-centric cost function, and gave bounds for various classical problems. Iacono (Section 3.10) proposed to model locality-of-reference more explicitly and compared his proposal to the external memory and cache-oblivious model. Sen (Section 3.27) proposed the novel paradigm HAIbrid augmenting classic data structures with artificial intelligence.

Final Thoughts

The organizers would like to thank the Dagstuhl team for their continuous support; the welcoming atmosphere made the seminar both highly productive and enjoyable. They also thank all participants for their contributions to this seminar.

2 Table of Contents

Executive Summary

Gerth Stølting Brodal, Ulrich Carsten Meyer, Markus E. Nebel, and Robert Sedgewick 104

Overview of Talks

| | |
|--|-----|
| Fragile Complexity of Comparison-based Algorithms <i>Peyman Afshani</i> | 108 |
| SPARQL Autocompletion – a Quick Intro and a Nice Open Problem <i>Hannah Bast</i> | 108 |
| Bloom Filters, Adaptivity, and the Dictionary Problem <i>Michael A. Bender</i> | 109 |
| Thrill: Pipelined External Memory Processing in the Cloud with C++ <i>Timo Bingmann</i> | 109 |
| Optimal Hashing in External Memory <i>Alexander Conway</i> | 110 |
| Constant-Time Retrieval with $O(\log m)$ Extra Bits <i>Martin Dietzfelbinger</i> | 110 |
| Dictionary Fragmentation, Affine IOs & File System Aging <i>Martin Farach-Colton</i> | 111 |
| Bidirectional Text Compression in External Memory <i>Johannes Fischer</i> | 111 |
| Polynomial Time Algorithms for Constructing Optimal AIFV Codes <i>Mordecai Golin</i> | 112 |
| The cache-oblivious model is better than you thought <i>John Iacono</i> | 112 |
| Lower Bounds for Oblivious Data Structures <i>Riko Jacob</i> | 113 |
| Modeling (and Other Research) Opportunities in Cloud and External Memory <i>Rob Johnson</i> | 113 |
| The Optimal (?) Complexity of Set-disjointness Data-structures <i>Tsvi Kopelowitz</i> | 114 |
| Enumerations Derived from Compact Encodings <i>Jérémie Lumbroso</i> | 114 |
| Affirmative Sampling <i>Conrado Martinez</i> | 115 |
| Fair Division of Indivisible Goods <i>Kurt Mehlhorn</i> | 115 |
| Reflections On Cost <i>Robert Muth</i> | 116 |
| Median-of- k Jumlists and Dangling-Min BSTs <i>Markus E. Nebel and Sebastian Wild</i> | 116 |

| | |
|---|-----|
| What We Learned About Dynamic GPU Data Structures <i>John D. Owens</i> | 117 |
| Parallel and I/O-efficient Randomisation of Massive Networks using Global Curve- ball Trades <i>Manuel Penschuck and Ulrich Carsten Meyer</i> | 117 |
| A Cactus-type Structure for Vertex Connectivity? <i>Seth Pettie</i> | 118 |
| An Algorithm for Real-Time Matching of Requests to Resources <i>Sharath Raghvendra</i> | 118 |
| What is the Cost of Staying in Perfect Shape (if You are a B-Tree)? <i>Rolf Fagerberg and Ulrich Carsten Meyer</i> | 119 |
| Persistent Data Structures for Full-Text Indexes <i>Alejandro Salinger</i> | 119 |
| Hashing with Linear Probing and Referential Integrity <i>Peter Sanders</i> | 120 |
| Cardinality Estimation: A Poster Child for Algorithm Science <i>Robert Sedgewick</i> | 120 |
| HAIbrid Algorithms <i>Siddhartha Sen</i> | 120 |
| External Memory Output Sensitive Triangle Enumeration <i>Francesco Silvestri</i> | 121 |
| Concurrent Connected Components <i>Robert Endre Tarjan</i> | 121 |
| On the Hardness and Inapproximability of Recognizing Wheeler Graphs <i>Sharma V. Thankachan</i> | 121 |
| Entropy Trees & Range-Minimum Queries in Optimal Average-Case Space <i>Sebastian Wild</i> | 122 |
| Mining Subgraphs from a Big Graph: Solution and Challenges <i>Da Yan</i> | 122 |
| Participants | 124 |

3 Overview of Talks

3.1 Fragile Complexity of Comparison-based Algorithms

Peyman Afshani (Aarhus University, DK)

License  Creative Commons BY 3.0 Unported license
 Peyman Afshani

Joint work of Peyman Afshani, Rolf Fagerberg, David Hammer, Riko Jacob, Irina Kostitsyna, Ulrich Meyer, Manuel Penschuck, Nodari Sitchinava

Main reference Peyman Afshani, Rolf Fagerberg, David Hammer, Riko Jacob, Irina Kostitsyna, Ulrich Meyer, Manuel Penschuck, Nodari Sitchinava: “Fragile Complexity of Comparison-Based Algorithms”, CoRR, Vol. abs/1901.02857, 2019.

URL <https://arxiv.org/abs/1901.02857>

We initiate a study of algorithms with a focus on the computational complexity of individual elements, and introduce the fragile complexity of comparison-based algorithms as the maximal number of comparisons any individual element takes part in. We give a number of upper and lower bounds on the fragile complexity for fundamental problems, including Minimum, Selection, Sorting and Heap Construction. The results include both deterministic and randomized upper and lower bounds, and demonstrate a separation between the two settings for a number of problems. The depth of a comparator network is a straight-forward upper bound on the worst case fragile complexity of the corresponding fragile algorithm. We prove that fragile complexity is a different and strictly easier property than the depth of comparator networks, in the sense that for some problems a fragile complexity equal to the best network depth can be achieved with less total work and that with randomization, even a lower fragile complexity is possible.

3.2 SPARQL Autocompletion – a Quick Intro and a Nice Open Problem

Hannah Bast (Universität Freiburg, DE)

License  Creative Commons BY 3.0 Unported license
 Hannah Bast

I gave a brief introduction to knowledge bases and Wikidata, the largest general-purpose knowledge base at the time of this talk (over 7 billion triples). I showed some demos of QLever, our query engine that supports SPARQL queries on knowledge bases as large and complex as Wikidata (SPARQL is the de facto standard query language for knowledge bases). In particular, I showed QLever’s nifty autocompletion feature and I formulated an interesting open problem that arises in the context of making this autocompletion more efficient for very large knowledge bases like Wikidata. The problem in a nutshell: given a set S of sets, compute a set P of “patterns” (frequent subsets of sets from S) such that sets from S can be written as the disjoint union of few such patterns. See the slides of the talk for details.

3.3 Bloom Filters, Adaptivity, and the Dictionary Problem

Michael A. Bender (Stony Brook University, US)

License © Creative Commons BY 3.0 Unported license
© Michael A. Bender

Joint work of Michael A. Bender, Martin Farach-Colton, Mayank Goswami, Rob Johnson, Sam McCauley, Shikha Singh

Main reference Michael A. Bender, Martin Farach-Colton, Mayank Goswami, Rob Johnson, Samuel McCauley, Shikha Singh: “Bloom Filters, Adaptivity, and the Dictionary Problem”, in Proc. of the 59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018, pp. 182–193, IEEE Computer Society, 2018.

URL <https://doi.org/10.1109/FOCS.2018.00026>

A Bloom filter (or alternative) maintains a compact, probabilistic representation of a set S of keys from a universe U . The price of being small is that there is a (bounded) probability of false positives.

This talk presets alternatives to Bloom filters that are faster, more space efficient, and support a wider range of operations. We show how these filters can adapt based on the results of past queries.

Joint work with Martin Farach-Colton, Mayank Goswami, Rob Johnson, Sam McCauley, and Shikha Singh.

3.4 Thrill: Pipelined External Memory Processing in the Cloud with C++

Timo Bingmann (KIT – Karlsruhe Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license
© Timo Bingmann

Joint work of Timo Bingmann, Simon Gog, Florian Kurpicz

Main reference Timo Bingmann, Simon Gog, Florian Kurpicz: “Scalable Construction of Text Indexes with Thrill”, in Proc. of the IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018, pp. 634–643, IEEE, 2018.

URL <https://doi.org/10.1109/BigData.2018.8622171>

We present on-going work on our distributed external Big Data processing framework Thrill [1]. It is a C++ framework consisting of a set of basic scalable algorithmic primitives like mapping, reducing and sorting, which can be combined into larger more complex algorithms, such as WordCount, PageRank, k -means clustering, and suffix sorting. These complex algorithms can then be run on very large inputs using a distributed computing cluster with external memory on each host.

We discuss how external memory is used in Thrill and the abstractions with which pipelined data processing loops are built. In this context we present current work on accelerating the distributed sorting operation with online splitter selection. As a case study of Thrill and its current external memory sorting implementation we present our freshly published results on distributed suffix array construction algorithms run on the Amazon EC2 Cloud.

References

- 1 Bingmann, Timo, et al. “Thrill: High-performance algorithmic distributed batch data processing with C++.” 2016 IEEE International Conference on Big Data (Big Data).

3.5 Optimal Hashing in External Memory

Alexander Conway (Rutgers University – Piscataway, US)

License  Creative Commons BY 3.0 Unported license
 © Alexander Conway

Hash tables are a ubiquitous class of dictionary data structures. However, standard hash table implementations do not translate well into the external memory model, because they do not incorporate locality for insertions. Iacono and Patrascu established an update/query tradeoff curve for external hash tables: a hash table that performs insertions in $\mathcal{O}(\lambda/B)$ amortized IOs require $\Omega(\log_\lambda N)$ expected IOs for queries, where N is the number of items that can be stored in the data structure, B is the size of a memory transfer, M is the size of memory, and λ is a tuning parameter. They provide a hashing data structure that meets this curve for λ that is $\Omega(\log \log M + \log_M N)$. We present a new and simpler optimal external memory hash table, the Bundle of Arrays Hash Table (BOA). BOAs are based on size-tiered LSMs and quotient filters, and are easy to implement. The BOA is optimal for a narrower range of λ . However, the simplicity of BOAs allows them to be readily modified to achieve the following results:

1. A new external memory data structure, the Bundle of Trees Hash Table (BOT), that matches the performance of the IP hash table, while retaining some of the simplicity of the BOAs.
2. The cache-oblivious Bundle of Trees Hash Table (COBOT), the first cache-oblivious hash table. This data structure matches the optimality of BOTs and IP hash tables over the same range of λ .

3.6 Constant-Time Retrieval with $\mathcal{O}(\log m)$ Extra Bits

Martin Dietzfelbinger (TU Ilmenau, DE)

License  Creative Commons BY 3.0 Unported license
 © Martin Dietzfelbinger

Joint work of Martin Dietzfelbinger, Stefan Walzer

Main reference Martin Dietzfelbinger, Stefan Walzer: “Constant-Time Retrieval with $\mathcal{O}(\log m)$ Extra Bits”, in Proc. of the 36th International Symposium on Theoretical Aspects of Computer Science, STACS 2019, March 13-16, 2019, Berlin, Germany, LIPIcs, Vol. 126, pp. 24:1–24:16, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2019.

URL <http://dx.doi.org/10.4230/LIPIcs.STACS.2019.24>

For a set U (the universe), retrieval is the following problem. Given a finite subset S of U of size m and $f: S \rightarrow \{0, 1\}^r$ for a small constant r , build a data structure D_f with the property that for a suitable query algorithm we have $\text{query}(D_f, x) = f(x)$ for all $x \in S$. For $x \in U \setminus S$ the value $\text{query}(D_f, x)$ is arbitrary in $\{0, 1\}^r$. The number of bits needed for D_f should be $(1 + \varepsilon)rm$ with overhead $\varepsilon = \varepsilon(m) \geq 0$ as small as possible, while the query time should be small. Of course, the time for constructing D_f is relevant as well. We assume fully random hash functions on U with constant evaluation time are available. It is known that with ε about 0.09 one can achieve linear construction time and constant query time, and with overhead $\varepsilon_k \sim e^{-k}$ it is possible to have $\mathcal{O}(k)$ query time and $\mathcal{O}(m^{1+\alpha})$ construction time, for arbitrary $\alpha > 0$. Furthermore, a theoretical construction with $\varepsilon = \mathcal{O}((\log \log m)/\sqrt{\log m})$ gives constant query time and linear construction time. Known constructions avoiding all overhead, except for a seed value of size $\mathcal{O}(\log \log m)$, require logarithmic query time. In this paper, we present a method for treating the retrieval problem with overhead $\varepsilon = \mathcal{O}((\log m)/m)$,

which corresponds to $\mathcal{O}(1)$ extra memory words ($\mathcal{O}(\log m)$ bits), and an extremely simple, constant-time query operation. The price to pay is a construction time of $\mathcal{O}(m^2)$. We employ the usual framework for retrieval data structures, where construction is effected by solving a sparse linear system of equations over the 2-element field F_2 and a query is effected by a dot product calculation. Our main technical contribution is the design and analysis of a new and natural family of sparse random linear systems with m equations and $(1 + \varepsilon)m$ variables, which combines good locality properties with high probability of having full rank. Paying a larger overhead of $\varepsilon = \mathcal{O}((\log m)/m^\alpha)$, the construction time can be reduced to $\mathcal{O}(m^{1+\alpha})$ for arbitrary constant $0 < \alpha < 1$. In combination with an adaptation of known techniques for solving sparse linear systems of equations, our approach leads to a highly practical algorithm for retrieval. In a particular benchmark with $m = 10^7$ we achieve an order-of-magnitude improvement over previous techniques with ε about 0.24% instead of the previously best result of ε about 3%, with better query time and no significant sacrifices in construction time.

3.7 Dictionary Fragmentation, Affine IOs & File System Aging

Martin Farach-Colton (Rutgers University – Piscataway, US)

License © Creative Commons BY 3.0 Unported license
© Martin Farach-Colton

B-trees fragment because nodes are small. Be-trees don't fragment because nodes are large. We test this hypothesis on file systems and find the modern file systems age by factors of 20 or greater, except for BetrFS, which does not age at all.

3.8 Bidirectional Text Compression in External Memory

Johannes Fischer (TU Dortmund, DE)

License © Creative Commons BY 3.0 Unported license
© Johannes Fischer

Joint work of Johannes Fischer, Patrick Dinklage, Jonas Ellert, Dominik Köppl, Manuel Penschuk

We present a new algorithm for text compression in external memory. It works by substituting repeated substrings by references, which can point either to the left or to the right of the current position (unlike Lempel-Ziv-77, which can only refer to the left). The algorithm is based on the “permuted longest common prefix array”, which can be computed efficiently while suffix-sorting the text, for which good external memory algorithms exist. We test our algorithms on inputs of up to 128 GiB on a computer with 16 GiB of RAM, where it uses much less space than the best external memory implementations of the LZ77-algorithm, while showing a similar compression rate.

3.9 Polynomial Time Algorithms for Constructing Optimal AIFV Codes

Mordecai Golin (HKUST – Kowloon, HK)

License © Creative Commons BY 3.0 Unported license
© Mordecai Golin

Joint work of Mordecai Golin, Elfarouk Harb

Main reference Mordecai J. Golin, Elfarouk Y. Harb: “Polynomial Time Algorithms for Constructing Optimal AIFV Codes”, in Proc. of the Data Compression Conference, DCC 2019, Snowbird, UT, USA, March 26-29, 2019, pp. 231–240, IEEE, 2019.

URL <http://dx.doi.org/10.1109/DCC.2019.00031>

Huffman Codes are *optimal* Fixed-to-Variable (FV) codes if every source symbol can only be encoded by one codeword. Relaxing these constraints permits constructing better FV codes. More specifically, recent work has shown that AIFV codes can beat Huffman coding. AIFV codes construct a set of different coding trees between which the code alternates and are only *almost instantaneous* (AI). This means that decoding a word might require a delay of a finite number of bits.

Current algorithms for constructing optimal AIFV codes are iterative processes. One iteration step improves the current set of trees to a “better” set. The process has been proven to finitely converge to the optimal code but with but no known bounds on the convergence time.

This paper derives a geometric interpretation of the space of AIFV codes. This permits the development of new polynomially time-bounded iterative procedures for constructing optimal AIFV codes.

For the simplest case we show that a binary search procedure can replace the current iterative process. For the more complicated cases we describe how to frame the problem as a linear programming problem with an exponential number of constraints but a polynomial time separability oracle. This permits using the Grötschel, Lovász and Schrijver ellipsoid method to solve the problem in a polynomial number of steps.

This work will be presented at DCC’19.

3.10 The cache-oblivious model is better than you thought

John Iacono (UL – Brussels, BE)

License © Creative Commons BY 3.0 Unported license
© John Iacono

Joint work of John Iacono, Varunkumar Jayapaul, Ben Karsin

Main reference John Iacono, Varunkumar Jayapaul, Ben Karsin: “Locality”, CoRR, Vol. abs/1902.07928, 2019.

URL <https://arxiv.org/abs/1902.07928>

The performance of modern computation is characterized by locality of reference, that is, it is cheaper to access data that has been accessed recently than a random piece of data. This is due to many architectural features including caches, lookahead, address translation and the physical properties of a hard disk drive; attempting to model all the components that constitute the performance of a modern machine is impossible, especially for general algorithm design purposes. What if one could prove an algorithm is asymptotically optimal on all systems that reward locality of reference, no matter how it manifests itself within reasonable limits? We show that this is possible, and that algorithms that are asymptotically optimal in the cache-oblivious model are asymptotically optimal in any reasonable locality-of-reference rewarding setting. This is surprising as the cache-oblivious model envisions a particular

architectural model involving blocked memory transfer into a multi-level hierarchy of caches of varying sizes, and was not designed to directly model locality-of-reference correlated performance.

3.11 Lower Bounds for Oblivious Data Structures

Riko Jacob (IT University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license
© Riko Jacob

Joint work of Riko Jacob, Kasper Green Larsen, Jesper Buus Nielsen

Main reference Riko Jacob, Kasper Green Larsen, Jesper Buus Nielsen: “Lower Bounds for Oblivious Data Structures”, in Proc. of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019, pp. 2439–2447, SIAM, 2019.

URL <https://doi.org/10.1137/1.9781611975482.149>

An oblivious data structure is a data structure where the memory access patterns reveals no information about the operations performed on it. Such data structures were introduced by Wang et al. [ACM SIGSAC’14] and are intended for situations where one wishes to store the data structure at an untrusted server. One way to obtain an oblivious data structure is simply to run a classic data structure on an oblivious RAM (ORAM). Until very recently, this resulted in an overhead of $\omega(\log n)$ for the most natural setting of parameters. Moreover, a recent lower bound for ORAMs by Larsen and Nielsen [CRYPTO’18] show that they always incur an overhead of at least $\Omega(\log n)$ if used in a black box manner. To circumvent the $\omega(\log n)$ overhead, researchers have instead studied classic data structure problems more directly and have obtained efficient solutions for many such problems such as stacks, queues, dequeues, priority queues and search trees. However, none of these data structures process operations faster than $\Theta(\log n)$, leaving open the question of whether even faster solutions exist. In this paper, we rule out this possibility by proving $\Omega(\log n)$ lower bounds for oblivious stacks, queues, dequeues, priority queues and search trees.

3.12 Modeling (and Other Research) Opportunities in Cloud and External Memory

Rob Johnson (VMware – Palo Alto, US)

License © Creative Commons BY 3.0 Unported license
© Rob Johnson

This talk will describe several trends in cloud computing that can serve as inspiration for new models and new algorithms and data structure research.

3.13 The Optimal (?) Complexity of Set-disjointness Data-structures

Tsvi Kopelowitz (Bar-Ilan University – Ramat Gan, IL)

License  Creative Commons BY 3.0 Unported license
© Tsvi Kopelowitz

In the set-disjointness problem the goal is to preprocess a family of sets $F = \{S_1, S_2, \dots, S_k\}$, all from universe U , so that given two sets $S_i, S_j \in F$, one can quickly establish whether the two sets are disjoint or not. If N is the sum of the sizes of the sets in F , then let N^p be the preprocessing time and let N^q be the query time. A folklore combinatorial algorithm has a tradeoff curve of $p + q = 2$, which is optimal (up to sub-polynomial terms) for combinatorial algorithms, assuming the combinatorial BMM conjecture. In SODA'16, Kopelowitz, Pettie, and Porat showed that, based on the 3SUM, there is a conditional lower bound curve of $p + 2q \geq 2$, and so there exists a gap between the upper bound curve and the lower bound curve when allowing non-combinatorial techniques.

In this talk we will show that both curves can be improved. Specifically, if one assumes that the constant in the exponent of fast matrix multiplication is $\omega = 2$, then one can obtain an upper bound curve of $p + 2q \geq 2$ for $q \leq 1/3$ (matching the 3SUM based lower bound for this case), and $2p + q = 3$ for $q \geq 1/3$. Moreover, we introduce a new conjecture on the time required for detecting a triangle in an unbalanced tripartite graph, which is closely related to the triangle detection conjecture for general graphs, and is used to show that the new upper bound curve for set-disjointness is tight.

3.14 Enumerations Derived from Compact Encodings

Jérémie Lumbroso (Princeton University, US)

License  Creative Commons BY 3.0 Unported license
© Jérémie Lumbroso

Joint work of Jérémie Lumbroso, Maryam Bahrani, Cédric Chauve, Éric Fusy, Jessica Shi
Main reference Jérémie O. Lumbroso, Jessica Shi: “Exponential Bounds on Graph Enumerations from Vertex Incremental Characterizations”, in Proc. of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2018, New Orleans, LA, USA, January 8-9, 2018., pp. 118–132, SIAM, 2018.

URL <http://dx.doi.org/10.1137/1.9781611975062.11>

We compare several methodologies to either compute or approximate the enumeration of a combinatorial class:

1. The “compact encoding” methodology, in which an encoding of the objects of the combinatorial class is designed, and then the enumeration sequence is bounded by lower/upper-bounding the number of bits required by the encoding for any object of a given size.
2. A method using the analytic combinatorics toolset, in which a bijection allows us to reduce the objects of the combinatorial class to an easier to count class, which can then be exactly enumerated using existing techniques.
3. A “hybrid” methodology, in which we design an encoding for the combinatorial class, but then count all possible encodings using analytic combinatorics tree enumeration techniques.

We argue that the third method provides a good trade-off between technical complexity and accuracy. We illustrate these arguments on various examples taken from recent work on the enumeration of unlabeled graph classes, including papers by Nakano *et al.* [2], Lumbroso and Shi [3], and Chauve *et al.* [1].

This is joint work with several researchers: Maryam Bahrani (Princeton), Cédric Chauve (Simon Fraser University), Éric Fusy (Polytechnique), Jessica Shi (MIT).

References

- 1 Chauve, Cédric, Éric Fusy, and Jérémie Lumbroso. *An Exact Enumeration of Distance-Hereditary Graphs*, in Proceedings of the Fourteenth Meeting on Analytic Algorithmics and Combinatorics (ANALCO 2017), C. Martínez and M. D. Ward, Eds., Barcelona, Spain: Society for Industrial and Applied Mathematics, pp. 31–45, Jan. 2017. doi: 10.1137/1.9781611974775.3.
- 2 Nakano, Shin-ichi, Ryuhei Uehara, and Takeaki Uno. *A new approach to graph recognition and applications to distance-hereditary graphs*, in Journal of computer science and technology 24.3, pp. 517–533, 2009.
- 3 Lumbroso, Jérémie, and Jessica Shi.. *Exponential Bounds on Graph Enumerations from Vertex Incremental Characterizations*, in Proceedings of the Fifteenth Meeting on Analytic Algorithmics and Combinatorics (ANALCO 2018), M. Nebel and S. Wagner, Eds., New Orleans, USA: Society for Industrial and Applied Mathematics, pp. 118–132, Jan. 2018.

3.15 Affirmative Sampling

Conrado Martínez (UPC – Barcelona, ES)

License © Creative Commons BY 3.0 Unported license
© Conrado Martínez

Joint work of Conrado Martínez, Jérémie Lumbroso

Affirmative Sampling is a practical and efficient novel algorithm to obtain random samples of distinct elements from a data stream, its most salient feature being that the size S of the sample will, on expectation, grow with the (unknown) number n of distinct elements in the data stream. As any distinct element has the same probability to be sampled, and the sample size is greater when the cardinality (when the “diversity”) is greater, the samples that Affirmative Sampling delivers are more representative and enable more accurate inferences than those produced by any scheme where the sample size is fixed a priori – hence, its name.

3.16 Fair Division of Indivisible Goods

Kurt Mehlhorn (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
© Kurt Mehlhorn

Joint work of Bhaskar Ray Chaudhury, Yun Kuen Cheung, Jugal Garg, Naveen Garg, Martin Hoefer, Kurt Mehlhorn

Main reference Bhaskar Ray Chaudhury, Yun Kuen Cheung, Jugal Garg, Naveen Garg, Martin Hoefer, Kurt Mehlhorn: “On Fair Division for Indivisible Items”, in Proc. of the 38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2018, December 11–13, 2018, Ahmedabad, India, LIPIcs, Vol. 122, pp. 25:1–25:17, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.

URL <http://dx.doi.org/10.4230/LIPIcs.FSTTCS.2018.25>

We consider the task of assigning indivisible goods to a set of agents in a fair manner. Our notion of fairness is Nash social welfare, i.e., the goal is to maximize the geometric mean of the utilities of the agents. Each good comes in multiple items or copies, and the utility of an agent diminishes as it receives more items of the same good. The utility of a bundle of items for an agent is the sum of the utilities of the items in the bundle. Each agent has

a utility cap beyond which he does not value additional items. We give a polynomial time approximation algorithm that maximizes Nash social welfare up to a factor of $e^{1/e} \approx 1.445$. The computed allocation is Pareto-optimal and approximates envy-freeness up to one item up to a factor of $2 + \varepsilon$.

3.17 Reflections On Cost

Robert Muth (Google – New York, US)

License  Creative Commons BY 3.0 Unported license
 Robert Muth

This informal talk is a practitioner's view of running algorithms in the cloud at Google.

We discuss various forms of cost metrics, including total cost of ownership (TCO). We point out how cloud needs differ from the classical space/time metrics, make some predictions about how certain cost models will change and point out opportunities for new research and how theory can inform practice.

3.18 Median-of- k Jumplists and Dangling-Min BSTs

Markus E. Nebel (Universität Bielefeld, DE) and Sebastian Wild (University of Waterloo, CA)

License  Creative Commons BY 3.0 Unported license
 Markus E. Nebel and Sebastian Wild
Joint work of Markus E. Nebel, Elisabeth Neumann, Sebastian Wild
Main reference Markus E. Nebel, Elisabeth Neumann, Sebastian Wild: “Median-of- k Jumplists”, CoRR, Vol. abs/1609.08513, 2016.
URL <http://arxiv.org/abs/1609.08513>

In this talk we discuss a variant of jumplists where the jump pointer's target is chosen as the median of a random sample of size k . We present a new search strategy called spine search which allows to skip some elements of the list which the traditional algorithm would have inspected. We prove a precise asymptotic for the average number of key comparisons performed by this new strategy which shows that larger values of k imply an improved lookup time for the jump lists tend to be more balanced. On the other hand, insertions and deletions get more costly with increasing k since the need for rebalancing gets worse. As a second improvement we analyse the effect of omitting jump-pointer for sublists of length at most w , proving that a constant fraction of pointer is saved while search costs only increase by a constant additive term.

3.19 What We Learned About Dynamic GPU Data Structures

John D. Owens (University of California, Davis, US)

License © Creative Commons BY 3.0 Unported license
© John D. Owens

Joint work of John Owens, Martin Farach-Colton, as well as many coauthors

In this talk I discuss four dynamic data structures we built for the GPU – log-structured merge trees, quotient filters, linked lists and hash tables built atop them, and B-trees. I discuss principles that we followed in building them and then what we learned, including lessons from mapping work to threads vs. warps, issues with contention, the use of the cache hierarchy, surprising insertion throughput results for the LSM vs. the B-tree, memory allocation, resizing, and semantics.

References

- 1 Muhammad A. Awad, Saman Ashkiani, Rob Johnson, Martín Farach-Colton, and John D. Owens. *Engineering a High-Performance GPU B-Tree*. In *Proceedings of the 24th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, PPOPP 2019, February 2019.
- 2 Saman Ashkiani, Martin Farach-Colton, and John D. Owens. *A Dynamic Hash Table for the GPU*. *Proceedings of the 31st IEEE International Parallel and Distributed Processing Symposium*, IPDPS 2018, pages 419–429. May 2018.
- 3 Saman Ashkiani, Shengren Li, Martin Farach-Colton, Nina Amenta, and John D. Owens. *GPU LSM: A Dynamic Dictionary Data Structure for the GPU*. In *Proceedings of the 31st IEEE International Parallel and Distributed Processing Symposium*, IPDPS 2018, pages 430–440. May 2018.
- 4 Afton Geil, Martin Farach-Colton, and John D. Owens. *Quotient Filters: Approximate Membership Queries on the GPU*. In *Proceedings of the 31st IEEE International Parallel and Distributed Processing Symposium*, IPDPS 2018, pages 451–462. May 2018.

3.20 Parallel and I/O-efficient Randomisation of Massive Networks using Global Curveball Trades

Manuel Penschuck (Goethe-Universität – Frankfurt a. M., DE) and Ulrich Carsten Meyer (Goethe-Universität – Frankfurt a. M., DE)

License © Creative Commons BY 3.0 Unported license
© Manuel Penschuck and Ulrich Carsten Meyer

Joint work of Corrie Jacobien Carstens, Michael Hamann, Ulrich Meyer, Manuel Penschuck, Hung Tran, Dorothea Wagner

Main reference Corrie Jacobien Carstens, Michael Hamann, Ulrich Meyer, Manuel Penschuck, Hung Tran, Dorothea Wagner: “Parallel and I/O-efficient Randomisation of Massive Networks using Global Curveball Trades”, in Proc. of the 26th Annual European Symposium on Algorithms, ESA 2018, August 20-22, 2018, Helsinki, Finland, LIPIcs, Vol. 112, pp. 11:1–11:15, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2018.

URL <http://dx.doi.org/10.4230/LIPIcs.ESA.2018.11>

Graph randomisation is a crucial task in the analysis and synthesis of networks. It is typically implemented as an edge switching process (ESMC) repeatedly swapping the nodes of random edge pairs while maintaining the degrees involved. We discuss EM-ES, an I/O-efficient edge switching algorithm for this setting.

In this context, Curveball is a novel approach that instead considers the whole neighbourhoods of randomly drawn node pairs. Its Markov chain converges to a uniform distribution,

and experiments suggest that it requires less steps than the established ESMC. Since trades however are more expensive, we study Curveball’s practical runtime by introducing the first efficient Curveball algorithms: the I/O-efficient EM-CB for simple undirected graphs and its internal memory pendant IM-CB.

Further, we investigate global trades processing every node in a graph during a single super step, and show that undirected global trades converge to a uniform distribution and perform superior in practice. We then discuss EM-GCB and EM-PGCB for global trades and give experimental evidence that EM-PGCB achieves the quality of the state-of-the-art ESMC algorithm EM-ES nearly one order of magnitude faster.

3.21 A Cactus-type Structure for Vertex Connectivity?

Seth Pettie (University of Michigan – Ann Arbor, US)

License  Creative Commons BY 3.0 Unported license
© Seth Pettie

It is well known that the set of all global minimum edge cuts in a graph can be succinctly represented as an $\mathcal{O}(n)$ -space “cactus tree”. In this talk I’ll survey efforts to develop an analogue of the cactus structure for k -vertex connectivity, such as the block tree ($k = 1$), the SPQR tree ($k = 2$), and the structure proposed by Cohen et al. for general k -connectivity.

3.22 An Algorithm for Real-Time Matching of Requests to Resources

Sharath Raghvendra (Virginia Polytechnic Institute – Blacksburg, US)

License  Creative Commons BY 3.0 Unported license
© Sharath Raghvendra

Main reference Sharath Raghvendra: “A Robust and Optimal Online Algorithm for Minimum Metric Bipartite Matching”, in Proc. of the Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016, September 7-9, 2016, Paris, France, LIPIcs, Vol. 60, pp. 18:1–18:16, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.

URL <http://dx.doi.org/10.4230/LIPIcs.APPROX-RANDOM.2016.18>

Main reference Krati Nayyar, Sharath Raghvendra: “An Input Sensitive Online Algorithm for the Metric Bipartite Matching Problem”, in Proc. of the 58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017, pp. 505–515, IEEE Computer Society, 2017.

URL <https://doi.org/10.1109/FOCS.2017.53>

In the online minimum-metric bipartite matching problem, we are given a set S of server locations in a metric space. The locations of requests are revealed one at a time and when a request is revealed, we must immediately and irrevocably match it to a “free” server. The cost of matching a server to request is given by the distance between the two locations (which we assume satisfies triangle inequality). The objective of this problem is to come up with a matching of servers to requests which is competitive with respect to the minimum-cost matching of S and R .

In this talk, I present an online algorithm that, in the adversarial request generation model, achieves near-optimal competitive ratio for any metric space. I will also show that this algorithm achieves an optimal competitive ratio of $2H_n - 1$ in the random arrivals model. I will show that this algorithm not only achieves near-optimal bounds in adversarial settings but also performs well in practical settings.

3.23 What is the Cost of Staying in Perfect Shape (if You are a B-Tree)?

Rolf Fagerberg (University of Southern Denmark – Odense, DK) and Ulrich Carsten Meyer (Goethe-Universität – Frankfurt a. M., DE)

License  Creative Commons BY 3.0 Unported license
© Rolf Fagerberg and Ulrich Carsten Meyer

Joint work of Rolf Fagerberg, David Hammer, Ulrich Carsten Meyer

Any B-tree has height at least $\lceil \log_B(n) \rceil$. Static B-trees achieving this are easy to build, but in the dynamic case, standard B-tree rebalancing algorithms only maintain a height within a constant factor of this optimum. We investigate exactly how close to $\lceil \log_B(n) \rceil$ the height of dynamic B-trees can be maintained and at what rebalancing costs. As usual, cost means the number of nodes accessed. We prove a lower bound on the cost of maintaining optimal height $\lceil \log_B(n) \rceil$, which shows that this cost must increase from $\Omega(1/B)$ to $\Omega(n/B)$ rebalancing per update as n grows from one power of B to the next. We also provide an almost matching upper bound, demonstrating this lower bound to be essentially tight. We then give a variant upper bound which can maintain near-optimal height at low cost. As two special cases, we can maintain optimal height for all but a vanishing fraction of values of n using $\Theta(\log_B(n))$ amortized rebalancing cost per update and we can maintain a height of optimal plus one using $O(1/B)$ amortized rebalancing cost per update. More generally, for any rebalancing budget, we can maintain (as n grows from one power of B to the next) optimal height essentially up to the point where the lower bound requires the budget to be exceeded, after which optimal height plus one is maintained. Finally, we prove that this balancing scheme gives B-trees with very good storage utilization.

3.24 Persistent Data Structures for Full-Text Indexes

Alejandro Salinger (SAP SE – Walldorf, DE)

License  Creative Commons BY 3.0 Unported license
© Alejandro Salinger

Joint work of Alejandro Salinger, Divya Venkatesan

We study an extension of the dynamic text collection problem in which versions of the text collection are maintained. In this problem, texts can be inserted to or deleted from any version of the collection, which in turn creates a new version of the collection. A solution should support the efficient location of a pattern in any version of the collection. An example of an application of this problem is in indexes of textual data in databases which support Multiversion Concurrency Control (MVCC). We present a full-text index which supports multi-version access while enabling users to modify any version of the indexed text collection. The index is based on the dynamic full-text indexes by Chan et al. [1] and Navarro and Mäkinen [2] and makes use of persistent data structures techniques to keep track of the versions of the collection in all data structures. We also present a simple variant of the above index which we use to compare its performance. It consists of a set of separate static indexes for the changes in the collection and achieves full persistence by maintaining a global version tree. We present an experimental evaluation of both indexes with respect to time and space usage for different operations.

This talk is based on the master's thesis of Divya Venkatesan (TU Kaiserslautern and SAP SE).

References

- 1 Chan, Ho-Leung and Hon, Wing-Kai and Lam, Tak-Wah and Sadakane, Kunihik. *Compressed Indexes for Dynamic Text Collections*. ACM Trans. Algorithms, 2007.
- 2 Mäkinen, Veli and Navarro, Gonzalo. *Dynamic Entropy-compressed Sequences and Full-text Indexes*. ACM Trans. Algorithms, 2008.

3.25 Hashing with Linear Probing and Referential Integrity

Peter Sanders (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license
© Peter Sanders

Main reference Peter Sanders: “Hashing with Linear Probing and Referential Integrity”, CoRR, Vol. abs/1808.04602, 2018.

URL <http://arxiv.org/abs/1808.04602>

We describe a variant of linear probing hash tables that never moves elements and thus supports referential integrity, i.e., pointers to elements remain valid while this element is in the hash table. This is achieved by the folklore method of marking some table entries as formerly occupied (tombstones). The innovation is that the number of tombstones is minimized. Experiments indicate that this allows an unbounded number of operations with bounded overhead compared to linear probing without tombstones (and without referential integrity).

3.26 Cardinality Estimation: A Poster Child for Algorithm Science

Robert Sedgewick (Princeton University, US)

License © Creative Commons BY 3.0 Unported license
© Robert Sedgewick

The problem of determining an approximate count of the number of items in a data stream has a rich history, dating back to a paper by Flajolet and Martin in the 1980s. This talk surveys the useful methods that have been invented, focusing on algorithms developed by Flajolet and co-authors that culminated in the HyperLogLog algorithm. These algorithms are characterized by careful analysis validated by experimentation, and are widely used today in cloud computing. The talk ends with a proposed new algorithm, HyperBitBit, that seems to provide accurate estimates for large streams using very little computation per item and only a very small amount of memory.

3.27 HAIbrid Algorithms

Siddhartha Sen (Microsoft – New York, US)

License © Creative Commons BY 3.0 Unported license
© Siddhartha Sen

HAIbrid (Human + AI) Algorithms synergize human solutions with AI to enhance the performance and adaptivity of hand-designed data structures and algorithms. These data structures and algorithms underlie our cloud storage, search, and scheduling systems. Rather

than avoiding AI or using it blindly, we seek the right combination – a new form of human-AI collaboration. I will describe a paradigm that combines reinforcement learning with the ability to ask counterfactual (“what if”) questions about any decision-making algorithm, provided there is sufficient randomness in its decisions. This paradigm can readily be applied to data structures like skip lists and treaps which are naturally randomized. Our ultimate goal is to create a “universal data structure” that delivers the best performance at every point in time, for any workload, through human + AI co-design.

3.28 External Memory Output Sensitive Triangle Enumeration

Francesco Silvestri (University of Padova, IT)

License © Creative Commons BY 3.0 Unported license
© Francesco Silvestri

Joint work of Rasmus Pagh, Francesco Silvestri

Triangle enumeration in the external memory model has been widely studied in a worst case scenario. It is not clear however if we can derive a more efficient algorithm whose I/O complexity is a function of the actual number of triangles of the input graph. In this talk, I provide a quick overview of the state-of-the-art, and describe some preliminary results on output-sensitive triangle enumeration in the external memory model.

3.29 Concurrent Connected Components

Robert Endre Tarjan (Princeton University, US)

License © Creative Commons BY 3.0 Unported license
© Robert Endre Tarjan

Joint work of Robert Endre Tarjan, Sixue (Cliff) Liu

Finding the connected components of a graph is one of the most basic graph problems. Although it is easy to find components sequentially using graph search or a disjoint set union algorithm, some important applications require finding the components of huge graphs, making sequential algorithms too slow. We describe recent progress on concurrent algorithms for this problem. Some simple algorithms seem surprisingly hard to analyze.

3.30 On the Hardness and Inapproximability of Recognizing Wheeler Graphs

Sharma V. Thankachan (University of Central Florida – Orlando, US)

License © Creative Commons BY 3.0 Unported license
© Sharma V. Thankachan

Joint work of Sharma V. Thankachan, Daniel Gibney

Main reference Daniel Gibney, Sharma V. Thankachan: “On the Hardness and Inapproximability of Recognizing Wheeler Graphs”, CoRR, Vol. abs/1902.01960, 2019.

URL <https://arxiv.org/abs/1902.01960>

In recent years the relationship between a newly defined class of graphs and several important string indexing structures has been discovered. This class of graphs, known as Wheeler graphs, were shown by Gagie et al. to model de Bruijn graphs, generalized compressed

suffix arrays, and several other BWT related structures. Moreover, the Wheeler graph axioms reveal a sufficient condition for a data structure to be indexed efficiently. In our work, we prove the NP-hardness of recognizing Wheeler graphs, in addition to providing an exponential time algorithm for the recognition problem which has better time complexity than the naive approach. We also show the APX-hardness of finding the minimum number of edges that must be removed to transform a graph into a Wheeler graph. On the other hand, we demonstrate that the dual of this optimization problem, finding the maximal Wheeler graph, admits a constant approximation.

3.31 Entropy Trees & Range-Minimum Queries in Optimal Average-Case Space

Sebastian Wild (University of Waterloo, CA)

License © Creative Commons BY 3.0 Unported license
© Sebastian Wild

Joint work of Munro, J. Ian, Nicholson, Patrick, Sebastian Wild

Main reference J. Ian Munro, Sebastian Wild: “Entropy Trees and Range-Minimum Queries In Optimal Average-Case Space”, CoRR, Vol. abs/1903.02533, 2019.

URL <https://arxiv.org/abs/1903.02533>

The range-minimum query (RMQ) problem is a fundamental data structuring task with numerous applications. Despite the fact that succinct solutions with worst-case optimal $2n + o(n)$ bits of space and constant query time are known, it has been unknown whether such a data structure can be made adaptive to the reduced entropy of random inputs (Davoodi et al. 2014). We construct a succinct data structure with the optimal $1.73n + o(n)$ bits of space on average for random RMQ instances, settling this open problem. Our solution relies on a compressed data structure for binary trees of independent interest.

References

- 1 Pooya Davoodi, Gonzalo Navarro, Rajeev Raman, and Srinivasa Rao Satti. *Encoding range minima and range top-2 queries*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372(2016):20130131–20130131, 2014. 10.1098/rsta.2013.0131

3.32 Mining Subgraphs from a Big Graph: Solution and Challenges

Da Yan (The University of Alabama – Birmingham, US)

License © Creative Commons BY 3.0 Unported license
© Da Yan

Joint work of Da Yan, Guimu Guo, Md Mashiur Rahman Chowdhury, M. Tamer özsu, John C.S. Lui, Weida Tan

Main reference Da Yan, Guimu Guo, Md Mashiur Rahman Chowdhury, M. Tamer Özsu, John C. S. Lui, Weida Tan: “T-thinker: a task-centric distributed framework for compute-intensive divide-and-conquer algorithms”, in Proc. of the 24th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPOPP 2019, Washington, DC, USA, February 16-20, 2019, pp. 411–412, ACM, 2019.

URL <https://doi.org/10.1145/3293883.3295709>

The problem of mining subgraphs from a big data graph finds numerous applications including social community detection and finding patterns from biological networks. Due to the high computational complexity of this problem, parallel and distributed computing is essential in order to scale to big graphs. However, existing cloud computing models such as MapReduce

and Google's Pregel are IO-bound and severely underutilize CPU cores when applied to this problem. This talk describes a new framework called G-thinker for large-scale graph mining, by briefly overviewing the parallel data structures used in G-thinker's design. G-thinker has been open-sourced at <http://www.cs.uab.edu/yanda/gthinker>. This talk will also bring some design challenges found during the development of G-thinker for open discussion.

Participants

- Peyman Afshani
Aarhus University, DK
- Hannah Bast
Universität Freiburg, DE
- Michael A. Bender
Stony Brook University, US
- Ioana Bercea
Tel Aviv University, IL
- Timo Bingmann
KIT – Karlsruher Institut für
Technologie, DE
- Gerth Stølting Brodal
Aarhus University, DK
- Alexander Conway
Rutgers University –
Piscataway, US
- Martin Dietzfelbinger
TU Ilmenau, DE
- Guy Even
Tel Aviv University, IL
- Tomer Even
Tel Aviv University, IL
- Rolf Fagerberg
University of Southern Denmark –
Odense, DK
- Martin Farach-Colton
Rutgers University –
Piscataway, US
- Johannes Fischer
TU Dortmund, DE
- Mordecai Golin
HKUST – Kowloon, HK
- Herman J. Haverkort
Universität Bonn, DE
- John Iacono
UL – Brussels, BE
- Riko Jacob
IT University of
Copenhagen, DK
- Rob Johnson
VMware – Palo Alto, US
- Tsvi Kopelowitz
Bar-Ilan University –
Ramat Gan, IL
- Moshe Lewenstein
Bar-Ilan University –
Ramat Gan, IL
- Jérémie Lumbroso
Princeton University, US
- Conrado Martinez
UPC – Barcelona, ES
- Kurt Mehlhorn
MPI für Informatik –
Saarbrücken, DE
- Ulrich Carsten Meyer
Goethe-Universität –
Frankfurt a. M., DE
- Friedhelm Meyer auf der Heide
Universität Paderborn, DE
- Ian Munro
University of Waterloo, CA
- Robert Muth
Google – New York, US
- Markus E. Nebel
Universität Bielefeld, DE
- John D. Owens
University of California at
Davis, US
- Manuel Penschuck
Goethe-Universität –
Frankfurt a. M., DE
- Seth Pettie
University of Michigan –
Ann Arbor, US
- Sharath Raghvendra
Virginia Polytechnic Institute –
Blacksburg, US
- Rajeev Raman
University of Leicester, GB
- Alejandro Salinger
SAP SE – Walldorf, DE
- Peter Sanders
KIT – Karlsruher Institut für
Technologie, DE
- Robert Sedgewick
Princeton University, US
- Siddhartha Sen
Microsoft – New York, US
- Francesco Silvestri
University of Padova, IT
- Robert Endre Tarjan
Princeton University, US
- Sharma V. Thankachan
University of Central Florida –
Orlando, US
- Sebastian Wild
University of Waterloo, CA
- Da Yan
The University of Alabama –
Birmingham, US
- Norbert Zeh
Dalhousie University –
Halifax, CA



Report from Dagstuhl Seminar 19052

Computational Methods for Melody and Voice Processing in Music Recordings

Edited by

Meinard Müller¹, Emilia Gómez², and Yi-Hsuan Yang³

1 Universität Erlangen-Nürnberg, DE, meinard.mueller@audiolabs-erlangen.de

2 UPF – Barcelona, ES, emilia.gomez@upf.edu

3 Academia Sinica – Taipei, TW, yang@citi.sinica.edu.tw

Abstract

In our daily lives, we are constantly surrounded by music, and we are deeply influenced by music. Making music together can create strong ties between people, while fostering communication and creativity. This is demonstrated, for example, by the large community of singers active in choirs or by the fact that music constitutes an important part of our cultural heritage. The availability of music in digital formats and its distribution over the world wide web has changed the way we consume, create, enjoy, explore, and interact with music. To cope with the increasing amount of digital music, one requires computational methods and tools that allow users to find, organize, analyze, and interact with music—topics that are central to the research field known as *Music Information Retrieval* (MIR). The Dagstuhl Seminar 19052 was devoted to a branch of MIR that is of particular importance: processing melodic voices (with a focus on singing voices) using computational methods. It is often the melody, a specific succession of musical tones, which constitutes the leading element in a piece of music. In the seminar we discussed how to detect, extract, and analyze melodic voices as they occur in recorded performances of a piece of music. Gathering researchers from different fields, we critically reviewed the state of the art of computational approaches to various MIR tasks related to melody processing including pitch estimation, source separation, singing voice analysis and synthesis, and performance analysis (timbre, intonation, expression). This triggered interdisciplinary discussions that leveraged insights from fields as disparate as audio processing, machine learning, music perception, music theory, and information retrieval. In particular, we discussed current challenges in academic and industrial research in view of the recent advances in deep learning and data-driven models. Furthermore, we explored novel applications of these technologies in music and multimedia retrieval, content creation, musicology, education, and human-computer interaction. 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, we present a more detailed overview of the participants' contributions (listed alphabetically by their last names) as well as of the ideas, results, and activities of the group meetings, the demo, and the music sessions.

Seminar January 27–February 1, 2019 – <http://www.dagstuhl.de/19052>

2012 ACM Subject Classification Information systems → Music retrieval, Applied computing → Sound and music computing

Keywords and phrases Acoustics of singing, audio signal processing, machine learning, music composition and performance, music information retrieval, music perception and cognition, music processing, singing voice processing, sound source separation, user interaction and interfaces

Digital Object Identifier 10.4230/DagRep.9.1.125

Edited in cooperation with Frank Zalkow



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

Computational Methods for Melody and Voice Processing in Music Recordings, *Dagstuhl Reports*, Vol. 9, Issue 1, pp. 125–177

Editors: Meinard Müller, Emilia Gómez, and Yi-Hsuan Yang



DAGSTUHL
REPORTS Dagstuhl Reports

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

1 Executive Summary

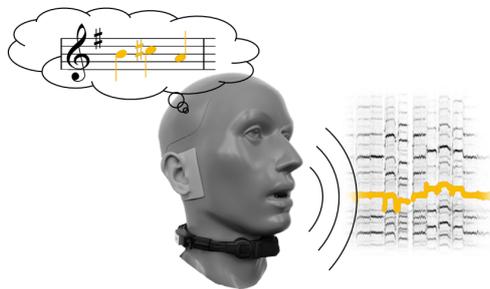
Meinard Müller (Universität Erlangen-Nürnberg, DE)

Emilia Gómez (UPF – Barcelona, ES)

Yi-Hsuan Yang (Academia Sinica – Taipei, TW)

License  Creative Commons BY 3.0 Unported license
© Meinard Müller, Emilia Gómez, and Yi-Hsuan Yang

In this executive summary, we give an overview of computational melody and voice processing and summarize the main topics covered in this seminar. We then describe the background of the seminar’s participants, the various activities, and the overall organization. Finally, we reflect on the most important aspects of this seminar and conclude with future implications and acknowledgments.



Overview

When asked to describe a specific piece of music, we are often able to sing or hum the main melody. In general terms, a *melody* may be defined as a linear succession of musical tones expressing a particular musical idea. Because of the special arrangement of tones, a melody is perceived as a coherent entity, which gets stuck in a listener’s head as the most memorable element of a song. As the original Greek term *melōidía* (meaning “singing” or “chanting”) implies, a melody is often performed by a human voice. Of course, a melody may also be played by other instruments such as a violin in a concerto or a saxophone in a jazz piece. Often, the melody constitutes the leading element in a composition, appearing in the foreground, while the accompaniment is in the background. Sometimes melody and accompaniment may even be played on a single instrument such as a guitar or a piano. Depending on the context and research discipline (e. g., music theory, cognition or engineering), one can find different descriptions of what may be meant by a melody. Most people would agree that the melody typically stands out in one way or another. For example, the melody often comprises the higher notes in a musical composition, while the accompaniment consists of the lower notes. Or the melody is played by some instrument with a characteristic timbre. In some performances, the notes of a melody may feature easily discernible time–frequency patterns such as vibrato, tremolo, or glissando. In particular, when considering performed music given in the form of audio signals, the detection, extraction, separation, and analysis of melodic voices becomes a challenging research area with many yet unsolved problems. In the following, we discuss some MIR tasks related to melody processing, indicating their relevance for fundamental research, commercial applications, and society.

The problem of detecting and separating melodic voices in music recordings is closely related to a research area commonly referred to as *source separation*. In general, audio signals are complex mixtures of different sound sources. The sound sources can be several people talking simultaneously in a room, different instruments playing together, or a speaker talking in the foreground with music being played in the background. The general goal of source separation is to decompose a complex sound mixture into its constituent components. Source separation methods often rely on specific assumptions such as the availability of multiple channels, where several microphones have been used to record the acoustic scene from different directions. Furthermore, the source signals to be identified are assumed to be independent in a statistical sense. In music, however, such assumptions are not applicable in many cases. For example, musical sound sources may outnumber the available information channels, such as a string quartet recorded in two-channel stereo. Also, sound sources in music are typically highly correlated in time and frequency. Instruments follow the same rhythmic patterns and play notes which are harmonically related. This makes the separation of musical voices from a polyphonic sound mixture an extremely difficult and generally intractable problem.

When decomposing a music signal, one strategy is to exploit music-specific properties and additional musical knowledge. In music, a source might correspond to a melody, a bass line, a drum track, or a general instrumental voice. The separation of the melodic voice, for example, may be simplified by exploiting the fact that the melody is often the leading voice, characterized by its dominant dynamics and by its temporal continuity. The track of a bass guitar may be extracted by explicitly looking at the lower part of the frequency spectrum. A human singing voice can often be distinguished from other musical sources due to characteristic time–frequency patterns such as vibrato. Besides such acoustic cues, score-informed source separation strategies make use of the availability of score representations to support the separation process. The score provides valuable information in two respects. On the one hand, pitch and timing of note events provide rough guidance within the separation process. On the other hand, the score offers a natural way to specify the target sources to be separated.

In this seminar, we discussed source separation techniques that are particularly suited for melodic voices. To get a better understanding of the problem, we approached source separation from different directions including model-based approaches that explicitly exploit acoustic and musical assumptions as well as data-driven machine learning approaches.

Given a music recording, melody extraction is often understood in the MIR field as the task of extracting a trajectory of frequency values that correspond to the pitch sequence of the dominant melodic voice. As said before, melody extraction and source separation are highly related: while melody extraction is much easier if the melodic source can be isolated first, the source separation process can be guided if the melodic pitch sequence is given a priori. However, both tasks have different goals and involve different challenges. The desired output of melody extraction is a trajectory of frequency values, which is often sufficient information for retrieval applications (e.g., query-by-humming or the search of a musical theme) and performance analysis. In contrast, for music editing and audio enhancement applications, source separation techniques are usually needed.

In the seminar, we addressed different problems that are related to melody extraction. For example, the melody is often performed by a solo instrument, which leads to a problem also known as *solo–accompaniment separation*. The estimation of the fundamental frequency of a quasi-periodic signal, termed *mono-pitch estimation*, is a long-studied problem with applications in speech processing. While mono-pitch estimation is now achievable with

reasonably high accuracy, the problem of *multi-pitch estimation* with the objective of estimating the fundamental frequencies of concurrent periodic sounds remains very challenging. This particularly holds for music signals, where concurrent notes stand in close harmonic relation. For extreme cases such as complex orchestral music where one has a high level of polyphony, multi-pitch estimation becomes intractable with today's methods.

Melodic voices are often performed by singers, and the singing voice is of particular importance in music. Humans use singing to create an identity, express their emotions, tell stories, exercise creativity, and connect while singing together. Because of its social, cultural, and educational impact, singing plays a central role in many parts of our lives, it has a positive effect on our health, and it creates a link between people, disciplines, and domains (e. g., music and language). Many people are active in choirs, and vocal music makes up an important part of our cultural heritage. In particular in Asian countries, karaoke has become a major cultural force performed by people of all age groups. Singing robots, vocaloids, or synthesizers such as Hatsune Miku¹ have made their way into the mass market in Japan. Thanks to digitization and technologies, the world wide web has become an important tool for amateur and professional singers to discover and study music, share their performances, get feedback, and engage with their audiences. An ever-increasing amount of music-related information is available to singers and singing enthusiasts, such as music scores² as well as audio and video recordings.³ Finally, music archives contain an increasing number of digitized audio collections of historic value from all around the world such as Flamenco music, Indian art music, Georgian vocal music, or Beijing Opera performances.

Due to its importance, we placed in our seminar a special emphasis on music technologies related to singing. This involves different research areas including singing analysis, description, and modeling (timbre, intonation, expression), singing voice synthesis and transformation, voice isolation/separation, and singing performance rating. Such research areas require a deep understanding of the way people produce and perceive vocal sounds. In our seminar, we discussed such issues with researchers having a background in singing acoustics and music performance.

Over the last years, as is also the case for other multimedia domains, many advances in music and audio processing have benefited from new developments in machine learning.

In particular, deep neural networks (DNNs) have found their way into MIR and are applied with increasing success to various MIR tasks including pitch estimation, melody extraction, sound source separation, and singing voice synthesis. The complex spectro-temporal patterns and relations found in music signals make this domain a challenging testbed for such new machine learning techniques. Music is different from many other types of multimedia. In a static image, for example, objects may occlude one another with the result that only certain parts are visible. In music, however, concurrent musical events may superimpose or blend each other in a more complicated way. Furthermore, as opposed to static images, music depends on time. Music is organized in a hierarchical way ranging from notes, bars, and motifs, to entire sections. As a result, one requires models that capture both short-term and long-term dependencies in music.

¹ https://en.wikipedia.org/wiki/Hatsune_Miku

² For example, the Choral Public Domain Library currently hosts free scores of at least 24963 choral and vocal works by at least 2820 composers, see <http://www.cpdl.org/>

³ See, for example, the material hosted at platforms such as YouTube or SoundCloud

In the seminar, we looked at the new research challenges that arise when designing music-oriented DNN architectures. Furthermore, considering the time-consuming and labor-intensive process of collecting human annotations of musical events and attributes (e.g., timbre, intonation, expression) in audio recordings, we addressed the issue of gathering large-scale annotated datasets that are needed for DNN-based approaches.

Participants and Group Composition

In our seminar, we had 32 participants, who came from various locations around the world including North America (4 participants from the U.S.), Asia (4 participants from Japan, 2 from Taiwan, 2 from Singapore, 1 from Korea, 1 from India), and Europe (18 participants from France, Germany, Netherlands, Spain, United Kingdom). More than half 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 field of music information retrieval, we also had participants with a background in musicology, acoustics, machine learning, signal processing, and other fields. By having experts working in technical as well as in non-technical disciplines, our seminar stimulated cross-disciplinary discussions, while highlighting opportunities for new collaborations among our attendees. Most of the participants had a strong musical background, some of them even having a dual career in an engineering discipline and music. This led to numerous social activities including singing and playing music together. In addition to geographical locations and research disciplines, we tried to foster variety in terms of seniority levels and presence of female researchers. In our seminar, 10 of the 32 participants were female, including three key researchers (Anja Volk, Emilia Gómez, and Johanna Devaney) from the “Women in Music Information Retrieval” (WiMIR)⁴ initiative.

In conclusion, by gathering internationally renowned scientists as well as younger promising researchers from different research areas, our seminar allowed us to gain a better understanding of the problems that arise when dealing with a highly interdisciplinary topic such as melody and voice processing—problems that cannot be addressed by simply using established research in signal processing or machine learning.

Overall Organization and Schedule

Dagstuhl seminars are known for having a high degree of flexibility and interactivity, which allows 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 tends to be on past research achievements. After the organizers gave an overview of the Dagstuhl concept and the seminar’s overall topic, we started the first day with self-introductions, where all participants introduced themselves and expressed their expectations and wishes for the seminar. We then continued with a small number of short (15 to 20 minutes) stimulus talks, where specific participants were asked to address some critical questions on melody and voice processing in a nontechnical fashion. Each of

⁴ <https://wimir.wordpress.com/>

these talks seamlessly moved towards an open discussion among all participants, where the respective presenters took over the role of a moderator. These discussions were well received and often lasted for more than half an hour. The first day closed with a brainstorming session on central topics covering the participants' interests while shaping the overall schedule and format for the next day. On the subsequent days, we continued having stimulus tasks interleaved with extensive discussions. Furthermore, we split into smaller groups, each group discussing a more specific topic in greater depth. The results and conclusions of these parallel group sessions, which lasted between 60 to 90 minutes, were then presented and discussed with the plenum. This mixture of presentation elements gave all participants the opportunity for presenting their ideas while avoiding a monotonous conference-like presentation format. On the last day, the seminar concluded with a session we called "self-introductions" where each participant presented his or her personal view on the seminar's results.

Additionally to the regular scientific program, we had several additional activities. First, we had a demo session on Thursday evening, where participants presented user interfaces, available datasets, and audio examples of synthesized singing voices. One particular highlight was the incorporation of singing practice in the seminar. In particular, we carried out a recording session on Wednesday afternoon, where we recorded solo and polyphonic singing performed by Dagstuhl participants. The goal of this recording session was to contribute to existing open datasets in the area of music processing. The singers were recorded with different microphone types such as throat and headset microphones to obtain clean recordings of the individual voices. All participants agreed that the recorded dataset should be made publicly available for research purposes. As preparation for these recordings, we assembled a choir consisting of ten to twelve amateur singers (all Dagstuhl participants) covering different voice sections (soprano, alto, tenor, bass). In the lunch breaks and the evening hours, the group met for regular rehearsals to practice different four-part choral pieces. These musical activities throughout the entire week not only supported the theoretical aspects of the seminar but also had a very positive influence on the group dynamics. Besides the recordings, we also had a concert on Thursday evening, where various participant-based ensembles performed a variety of music including classical music and folk songs.

Conclusions and Acknowledgment

Having a Dagstuhl seminar, we gathered researchers from different fields including information retrieval, signal processing, musicology, and acoustics. This allowed us to approach the problem of melody and voice processing by looking at a broad spectrum of data analysis techniques (including signal processing, machine learning, probabilistic models, user studies), by considering different domains (including text, symbolic, image, audio representations), and by drawing inspiration from the creative perspectives of the agents (composer, performer, listener) involved. As a key result of this seminar, we achieved some substantial progress towards understanding, modeling, representing, and extracting melody- and voice-related information using computational means.

The Dagstuhl seminar gave us the opportunity for having interdisciplinary discussions in an inspiring and retreat-like atmosphere. The generation of novel, technically oriented scientific contributions was not the main focus of the seminar. Naturally, many of the contributions and discussions were on a conceptual level, laying the foundations for future projects and collaborations. Thus, the main impact of the seminar is likely to take place in the medium and long term. Some more immediate results, such as plans to share research

data and software, also arose from the discussions. In particular, we plan to make the dataset recorded during the Dagstuhl seminar available to the research community. As further 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. Many of our participants were visiting Dagstuhl for the first time and enthusiastically praised the open and inspiring setting. The group dynamics were excellent with many personal exchanges and common activities. Some scientists expressed their appreciation for having the opportunity for prolonged discussions with researchers from neighboring research fields—something that is often impossible during conference-like events.

In conclusion, our expectations for the seminar were not only met but exceeded, in particular concerning networking and community building. We want to express our gratitude to the Dagstuhl board for giving us the opportunity to organize this seminar, the Dagstuhl office for their exceptional support in the organization process, and the entire Dagstuhl staff for their excellent service during the seminar. In particular, we want to thank Susanne Bach-Bernhard, Annette Beyer, Michael Gerke, and Michael Wagner for their assistance during the preparation and organization of the seminar.

2 Table of Contents

Executive Summary

Meinard Müller, Emilia Gómez, and Yi-Hsuan Yang 126

Stimulus Talks and Further Topics

Challenges in Melodic Similarity

Rachel Bittner 135

Singing Voice, Speech, or Something in Between

Estefanía Cano Cerón 135

Frequency Measurements and Perceived Pitch in Vocal Productions

Michèle Castellengo 136

Generative Models for Singing Voice

Pritish Chandna 137

Measuring Interdependence in Unison Choral Singing

Helena Cuesta 137

Can Listening Tests Help Understanding and Improving Data Models for Vocal Performance?

Johanna Devaney 138

Measuring and Modelling Intonation and Temperaments

Simon Dixon 138

What Makes Singing Unique?

Zhiyao Duan 140

From Science to Engineering to Deep Learning in Singing Processing

Emilia Gómez 141

Singing Information Processing

Masataka Goto 142

Melody and Voice Processing – Some Thoughts on the Seminar

Frank Kurth 143

Try, Try, Try Again: Rehearsals as a Data Source

Cynthia Liem 143

Singing Voice Separation: Recent Breakthroughs with Data-Driven Methods

Antoine Liutkus 144

Interactive Interfaces for Choir Rehearsal Scenarios

Meinard Müller, Sebastian Rosenzweig, and Frank Zalkow 145

Singing Interfaces and Visualizations

Tomoyasu Nakano 146

Social Voice Processing

Juhan Nam 146

Automatic Singing Transcription for Music Audio Signals

Ryo Nishikimi 147

Dominant Melody Estimation and Singing Voice Separation

Geoffroy Peeters 148

| | |
|--|-----|
| Russian/Ukrainian Traditional Polyphony: Musicological Research Questions and Challenges for MIR <i>Polina Proutskova</i> | 148 |
| Aspects of Melodic Similarity <i>Preeti Rao</i> | 149 |
| Extraction Techniques for Harmonic and Melodic Interval Analysis of Georgian Vocal Music <i>Sebastian Rosenzweig, Meinard Müller, and Frank Scherbaum</i> | 150 |
| Some Uncomfortable Statements about Melody Extraction <i>Justin Salamon</i> | 152 |
| Computational Analysis of Traditional Georgian Vocal Music <i>Frank Scherbaum, Nana Mzhavanadze, Sebastian Rosenzweig, Daniel Vollmer, Vlora Arifi-Müller, and Meinard Müller</i> | 153 |
| Measuring Intonation via Dissonance in Polyphonic Choral Singing <i>Sebastian J. Schlecht, Christof Weiß, Sebastian Rosenzweig, and Meinard Müller</i> . | 154 |
| Signal Processing for Multiple Fundamental Frequency Estimation <i>Li Su</i> | 155 |
| Augmented Vocal Production towards New Singing Style Development <i>Tomoki Toda</i> | 157 |
| Useless Evaluation versus User-less Evaluation <i>Julián Urbano</i> | 158 |
| Computational Modeling of Melody <i>Anja Volk</i> | 158 |
| Singing Voice Modelling for Language Learning <i>Ye Wang</i> | 160 |
| Analyzing and Visualizing Intonation in Polyphonic Choral Singing <i>Christof Weiß, Sebastian J. Schlecht, Sebastian Rosenzweig, and Meinard Müller</i> . | 161 |
| Melody Extraction for Melody Generation <i>Yi-Hsuan Yang</i> | 162 |
| Finding Musical Themes in Western Classical Music Recordings <i>Frank Zalkow, Stefan Balke, and Meinard Müller</i> | 164 |

Working Groups

| | |
|---|-----|
| Data Sets: What is Missing and How do We Get It? <i>Participants of Dagstuhl Seminar 19052</i> | 165 |
| Deep Learning Versus Acoustic Models <i>Participants of Dagstuhl Seminar 19052</i> | 165 |
| Insights <i>Participants of Dagstuhl Seminar 19052</i> | 166 |
| Singing Assessment and Performance Evaluation <i>Participants of Dagstuhl Seminar 19052</i> | 167 |

| | |
|--|-----|
| Subjective Evaluation and Objective Metrics for Singing Voice Generation and Separation | |
| <i>Participants of Dagstuhl Seminar 19052</i> | 169 |
| Symbolic Domain Melody Estimation | |
| <i>Participants of Dagstuhl Seminar 19052</i> | 170 |
| Transcription beyond Western Music | |
| <i>Participants of Dagstuhl Seminar 19052</i> | 171 |
| Demo Session | |
| <i>Participants of Dagstuhl Seminar 19052</i> | 173 |
| Music and Recording Sessions | |
| Choir Rehearsals | |
| <i>Christof Weiß, Sebastian Rosenzweig, Helena Cuesta, Frank Scherbaum, Emilia Gómez, and Meinard Müller</i> | 174 |
| Recording Documentation | |
| <i>Sebastian Rosenzweig, Helena Cuesta, Frank Scherbaum, Christof Weiß, Emilia Gómez, and Meinard Müller</i> | 175 |
| Participants | 177 |

3 Stimulus Talks and Further Topics

3.1 Challenges in Melodic Similarity

Rachel Bittner (Spotify – New York, US)

License © Creative Commons BY 3.0 Unported license
© Rachel Bittner

Many methods exist that generate reasonably accurate melody estimates from polyphonic music, particularly for vocal melodies. Looking beyond the problem of estimation itself, how do we measure the similarity between two estimated melodic sequences? In MIR we have most often chosen to estimate melody as a sequence of f_0 curves. However, the vast majority of work on melodic similarity has been done in the symbolic domain and focused on perceptual similarity. Work on query-by-humming in some sense measures melodic similarity, but focuses on shorter excerpts in a retrieval setting. This raises a number of questions:

- Should melodic similarity be computed in the symbolic domain? If so, what is the state of f_0 -to-note-conversion algorithms?
- Can we compute melodic similarity directly in the f_0 domain? How do we enforce robustness to differences in pitch curve characteristics and estimation errors?
- How do we build upon and incorporate the previous work on perceptual melodic similarity?

3.2 Singing Voice, Speech, or Something in Between

*Estefanía Cano Cerón (Fraunhofer IDMT, DE & A*STAR – Singapore, SG)*

License © Creative Commons BY 3.0 Unported license
© Estefanía Cano Cerón

Joint work of Estefanía Cano Cerón, Antonio Escamilla, Gustavo López Gil, Fernando Mora Ángel, José Ricardo Zapata

In the context of the ACMus research project, we are investigating automatic techniques for annotation and segmentation of digital archives of non-western music. In particular, we are focusing on a collection of traditional Colombian music compiled in the Músicas Regionales archive at the Universidad de Antioquia in Medellín, Colombia. Of particular interest are a series of recordings of vocal expressions of indigenous cultures native to different regions of the country. These vocal expressions can either be very close to speech (almost as reciting something), can include some melodic elements, or can be closer to the concept of singing voice from a Western music perspective (exhibiting a defined melodic line). From an automatic classification point of view, the traditional binary discrimination between speech and singing voice falls short, calling for a more general characterization of vocal expressions. For example, one may use a range $[0, 1]$ with 0 being pure speech, 1 being singing voice (from a Western music perspective), and allowing everything that falls in between these bounds. In this context, interesting research questions arise: Is the degree of melodic elements in these vocal expressions informative for the region where a recording was made? Can we conclude the functional aspects of these recordings (ritual, prayer, social, playful, healing) based on this characterization? Could these categories be shared between different cultures?

3.3 Frequency Measurements and Perceived Pitch in Vocal Productions

Michèle Castellengo (Sorbonne University – Paris, FR)

License  Creative Commons BY 3.0 Unported license
© Michèle Castellengo

The research developed in the MIR community aims to automatically extract the fundamental frequency of sounds and match it with a notated musical score. In general, it is assumed that the perceived pitch matches the measured frequency. In the case of singing, this postulate raises several questions related to the acoustic characteristics of the vocal source (such as unstable emission; rich and complex harmonic), the cognitive criteria of the “human voice” category (male or female, vowel listening), as well as the mental scale of reference of the listener (in this case the equal temperament). In my presentation, I focused on three examples that illustrate the importance of interactions between acoustic, psychophysical, and cognitive aspects involved in the interpretation of results.

- Perceived pitch accuracy of bel canto vocal techniques (vibrato versus trill, pitch of very short vibrated notes), see [1].
- Competition between the perception of pitch due to the periodicity of the signal and the spectral pitches coming from the vowel formants (octavian singing, diphonic singing), see [2].
- Perceptual emergence of a voice without a real existence (the Sardinian quintina), see [3].

Furthermore, I discussed issues related to traditional music and interval measurement. The confrontation with traditional polyphonies questions the unconscious mental references that condition our musical listening. In particular measuring intervals in cents, if perfectly adapted to piano music, can mislead the researcher when it comes to natural intervals. As an example, one can find in [4, pp. 403–404] the interval of a natural minor third (315.61 cents) in a song from Cameroon. Could we imagine a spectral analysis method based on the search for the harmonics common to both sounds of an interval [4, pp. 55]? See also [4, pp. 418–420] for analysis of stable multiphonic sounds.

References

- 1 Christophe d’Alessandro and Michèle Castellengo. *The Pitch of Short-Duration Vibrato Tones*. Journal of the Acoustical Society of America, 95 (3), 1994, pp. 1617–1630.
- 2 Michèle Castellengo and Nathalie Henrich Bernardoni. *Interplay Between Harmonics and Formants in Singing: When Vowels Become Music*. Proceedings of the International Symposium on Musical Acoustics (ISMA), Le Mans, France, 2014, pp. 433–438.
- 3 Michèle Castellengo, Bernard Lortat-Jacob, and Gilles Léothaud. *Pitch Perception: Five Voices with Four Sardinian Singers*. Proceedings of the International Symposium on Musical Acoustics (ISMA), Perugia, Italy, 2001, pp. 351–354.
- 4 Michèle Castellengo. *Écoute Musicale et Acoustique*. Eyrolles, Paris, France, 2015.

3.4 Generative Models for Singing Voice

Pritish Chandna (UPF – Barcelona, ES)

License © Creative Commons BY 3.0 Unported license
© Pritish Chandna

In the demo session of the Dagstuhl seminar, I presented a novel methodology for the synthesis of expressive singing voices using the NUS-48E corpus [1]. The approach is based on a generative adversarial network (GAN) for synthesizing the singing voice [2]. Furthermore, I discussed strategies for evaluating the quality of the synthesized results, see also the discussion of the working group on “Subjective Evaluation and Objectives Metrics for Singing Voice Generation and Separation” (Section 4.5).

References

- 1 Zhiyan Duan, Haotian Fang, Bo Li, Khe Chai Sim, and Ye Wang. *The NUS Sung and Spoken Lyrics Corpus: A Quantitative Comparison of Singing and Speech*. Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, Taiwan, 2013, pp. 1–9.
- 2 Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. *Generative Adversarial Networks*. Proceedings of the International Conference on Neural Information Processing Systems, Montreal, Canada, 2014, pp. 2672–2680.

3.5 Measuring Interdependence in Unison Choral Singing

Helena Cuesta (UPF – Barcelona, ES)

License © Creative Commons BY 3.0 Unported license
© Helena Cuesta

Choral singing is probably the most widespread type of singing [1], but there is still little research on the topic, especially from a computational perspective. In a choir, singers are influenced by other singers’ performances in terms of pitch and timing. In [3], we investigate the synchronization between singers by analyzing fundamental frequency (f_0) envelopes using two different features: the derivative of the f_0 curves [2] and the deviation from the target pitch specified by the score. In the first case, synchronization is assumed to be linear and measured using the Pearson correlation coefficient; in the second case, we consider the mutual information measure. Results suggest that the mutual information is a better metric for this task because of its non-linear nature; however, the question of which is the most suitable metric to measure this aspect of ensemble singing remains open for discussion.

References

- 1 Johan Sundberg. *The Science of the Singing Voice*. Northern Illinois University Press, 1987.
- 2 Helena Cuesta, Emilia Gómez, Agustín Martorell, and Felipe Loáiciga. *Analysis of Intonation in Unison Choir Singing*. Proceedings of the International Conference on Music Perception and Cognition, 2018. <http://doi.org/10.5281/zenodo.1319597>.
- 3 Helena Cuesta and Emilia Gómez. *Measuring Interdependence in Unison Choir Singing*. Late Breaking Demo of the International Conference on Music Information Retrieval (ISMIR), Paris, France, 2018.

3.6 Can Listening Tests Help Understanding and Improving Data Models for Vocal Performance?

Johanna Devaney (Brooklyn College, US)

License  Creative Commons BY 3.0 Unported license
© Johanna Devaney

The flexibility of the singing voice affords its great expressivity, but this flexibility also makes it challenging to model it computationally. In my presentation, I described an experiment on singer identification based on note-level descriptors (related to pitch, timing, loudness, and timbre) that compared the results of an SVM-based computational model and a listening experiment [1]. In this experiment, I considered not only the results but also the general implications for singing research and music information retrieval research in general.

References

- 1 Johanna Devaney. *Inter- Versus Intra-Singer Similarity and Variation in Vocal Performance*. *Journal of New Music Research*. 45(3), 2016, pp. 252–264.

3.7 Measuring and Modelling Intonation and Temperaments

Simon Dixon (Queen Mary University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Simon Dixon

Joint work of Simon Dixon, Matthias Mauch, Dan Tidhar, Klaus Frieler, Emmanouil Benetos, Tillman Weyde, Jiajie Dai

Melody and voice are themes running through several current research projects. In the context of the Trans-Atlantic Platform Digging into Data Challenge project “Dig that Lick: Analysing Large-scale Data for Melodic Patterns in Jazz Performances,” we are investigating the use of melodic patterns in jazz improvisation in a large corpus spanning a significant proportion of the recorded history of jazz. The research involves automatic extraction of the main melody voice [1], the recognition of musical structures, and their linkage to historical and social metadata [2]. This project informs the study of the transmission of musical ideas—in this instance “licks” and patterns, but more generally musical style [3]—across time and location.

This leads to another strand of research, on the singing voice, which began with a focus on intonation and the effects of musical context on singers’ ability to sing in tune [4, 5, 6, 7]. Wider questions are now being addressed to aid the study of performance and style, including the development of suitable features and representations for singing which capture and allow the comparison of continuous pitch trajectories, their segmentation, the use of dynamics, articulation, timbre (including vowel sounds and phonation modes [8]). Supported by the EU H2020 project “New Frontiers in Music Information Processing,” we are collaborating with DoReMIR Music Research on singing transcription for their ScoreCloud transcription service⁵. We are also developing source separation methods for voice [9], which will enable the analysis of accompanied singing.

Based on my previous work, I presented several studies on pitch, intonation, and singer interaction. Pitch is perhaps the most essential characteristic of musical sounds, being

⁵ www.scorecloud.com

the basis of both melody and harmony. Western music theory provides a framework for understanding pitch relationships in terms of intervals, scales and chords, expressed as a first approximation in indivisible units of semitones. Common music notation reflects this world-view. At the same time, it has been recognized since the time of Pythagoras that it is not possible for all theoretically consonant intervals to be perfectly “in tune”, and this has led to many theoretical and practical approaches to intonation, the realization of pitch, in music performance.

I presented investigations of intonation at two extremes of musical practice: a fixed-pitch instrument, the harpsichord, where the tuner determines the intonation of each pitch before the performance, and a variable-pitch instrument, the human voice, which can adjust the pitch of each note to the musical context and vary the pitch over the note’s duration. In each case, we have developed software tools for (semi-)automatic analysis of the pitch content from audio recordings.

We analyzed a collection of solo harpsichord CDs, estimated the inharmonicity and temperament of the harpsichord for each movement, and compared the measured temperaments with those given in the CD sleeve notes. The observed differences illustrate the tension between temperament as a theoretical construct and as a practical issue for professional performers and tuners. We conclude that “ground truth” is not always scientific truth and that content-based analysis has an essential role in the study of historical performance practice.

The second study investigates intonation and intonation drift in unaccompanied solo singing and proposes a simple intonation memory model that accounts for many of the effects observed. Singing experiments were conducted with 24 singers of varying ability. Over the duration of the recordings, approximately 50 seconds, a median absolute intonation drift of 11 cents was observed, which was smaller than the median note error (19 cents) but was significant in 22 percent of the recordings. Drift magnitude did not correlate with other measures of singing accuracy or singing experience. Neither a static intonation memory model nor a memoryless interval-based intonation model can account for the accuracy and drift behavior observed. The proposed causal model provides a better fit.

The third study looked at how pitch is negotiated by imperfect unaccompanied singers when they sing in pairs and the factors that influence pitch accuracy. Two singing conditions (unison versus 2-part harmony) were compared, along with an experimental condition varied which singers could hear their partners, measured in terms of pitch and interval errors. We found the following:

- Unison singing is more accurate than singing harmony.
- Singing solo is more accurate than singing with a partner.
- Singers adjust pitch to mitigate their partner’s error and preserve harmonic intervals at the expense of melodic intervals and absolute pitch.
- Other factors influencing pitch accuracy include score pitch, score harmonic interval, score melodic interval, musical background, vocal part, and individual differences.

References

- 1 Dogac Basaran, Slim Essid and Geoffroy Peeters. *Main Melody Extraction with Source-filter NMF and C-RNN*. International Society for Music Information Retrieval Conference, Paris, France, 2018, pp. 82–89.
- 2 Klaus Frieler, Frank Hoeger, Martin Pfeiderer and Simon Dixon. *Two Web Applications for Exploring Melodic Patterns in Jazz Solos*. International Society for Music Information Retrieval Conference, Paris, France, 2018, pp. 777–783.

- 3 Maria Panteli, Rachel Bittner, Juan Pablo Bello and Simon Dixon. *Towards the Characterization of Singing Styles in World Music*. IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, USA, 2017, pp. 636–640.
- 4 Matthias Mauch, Klaus Frieler and Simon Dixon. *Intonation in Unaccompanied Singing: Accuracy, Drift and a Model of Reference Pitch Memory*. Journal of the Acoustical Society of America, 136 (1), 2014, pp. 401–411.
- 5 Jiajie Dai and Simon Dixon. *Analysis of Vocal Imitations of Pitch Trajectories*. International Society for Music Information Retrieval Conference, New York, USA, 2016, pp. 87–93.
- 6 Jiajie Dai and Simon Dixon. *Analysis of Interactive Intonation in Unaccompanied SATB Ensembles*. International Society for Music Information Retrieval Conference, Suzhou, China, 2017, pp. 599–605.
- 7 Jiajie Dai and Simon Dixon. *Pitch Accuracy and Interaction in Unaccompanied Duet Singing*. Journal of the Acoustical Society of America, 145 (2), 2019, pp. 663–675.
- 8 Daniel Stoller and Simon Dixon. *Analysis and Classification of Phonation Modes in Singing*. International Society for Music Information Retrieval Conference, New York, USA, 2016, pp. 80–86.
- 9 Daniel Stoller, Sebastian Ewert and Simon Dixon. *Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation*. International Society for Music Information Retrieval Conference, Paris, France, 2018, pp. 334–340.
- 10 Simon Dixon, Matthias Mauch and Dan Tidhar. *Estimation of Harpsichord Inharmonicity and Temperament from Musical Recordings*. Journal of the Acoustical Society of America, 131 (1), 2012, pp. 878–887.
- 11 Dan Tidhar, Simon Dixon, Emmanouil Benetos and Tillman Weyde. *The Temperament Police*. Early Music, 42 (4), 2014, pp. 579–590.

3.8 What Makes Singing Unique?

Zhiyao Duan (University of Rochester, US)

License  Creative Commons BY 3.0 Unported license
© Zhiyao Duan

Singing is arguably the most popular kind of music throughout human history. People love singing or listening to singing on various occasions, and numerous styles of singing music have been composed, performed, recorded, and enjoyed. The reasons for this popularity, in my opinion, are manifold. First, singing is a musical behavior that is easier to learn and to perform compared to instrumental music. Second, singing often presents lyrics and is an enriched form of storytelling, which is at the core of human civilization. Third, singing voices generally show more flexibilities in pitch, dynamics, and timbre than musical instruments, which greatly enhance their expressiveness. Fourth, compared to musicians playing an instrument, singers have more freedom in expressing their emotions through facial expressions and body gestures, which significantly helps engage audiences.

Research on computational models for singing voices, in my opinion, needs to consider these unique properties of singing voices. For example, melody generation for singing voices may need to consider matching certain rhythms and tones of the natural speech of the underlying lyrics. This match is quite common in opera and folk songs of tonal languages (e. g., Chinese). Also, melody transcription may need to go beyond the standard piano-roll notation for instrumental music due to the extensive use of pitch glides and microtonality.

Raw pitch contours, however, seem to not bear enough abstraction for high-level processing. Some symbolic representations in between might be ideal. Furthermore, the visual aspects of singing, which have been neglected in the MIR community for a long time, may be crucial for singing analysis and synthesis. Based on these thoughts, I propose the following two novel research directions.

- **Lyrics-Informed Melody Generation for Chinese Opera.** Given lyrics (e. g., a poem or short story) and its speech intonation (e. g., Mandarin, Cantonese, other dialects), I propose to generate a melody for the lyrics automatically. The pitch contour of the melody should match with certain aspects of that of the speech intonation. It is an interesting question to investigate through a large corpora study which aspects should be considered for the match and which aspects allow for more flexibility.
- **Audio–Visual Analysis of Singing.** I propose to analyze the emotion and expressiveness of singing from the audio and visual modalities jointly. In particular, facial expressions and body movements should be analyzed and correlated with the audio signals. A related problem is audio–visual singing voice separation. Visual signals, especially the lip movements, provide cues about the singing activity and content, although some cues can be ambiguous or misleading depending on the singing style.

3.9 From Science to Engineering to Deep Learning in Singing Processing

Emilia Gómez (UPF – Barcelona, ES)

License © Creative Commons BY 3.0 Unported license
© Emilia Gómez

Joint work of Emilia Gómez, Merlijn Blaauw, Jordi Bonada, Pritisy Chadna and Helena Cuesta

Main reference Emilia Gómez, Merlijn Blaauw, Jordi Bonada, Pritisy Chadna, Helena Cuesta: “Deep Learning for Singing Processing: Achievements, Challenges and Impact on Singers and Listeners”. Keynote speech, 2018 Joint Workshop on Machine Learning for Music. The Federated Artificial Intelligence Meeting (FAIM), a joint workshop program of ICML, IJCAI/ECAI, and AAMAS, 2018.

URL <https://arxiv.org/abs/1807.03046>

The singing voice is probably the most complex “instrument” to analyze and synthesize, and a vast amount of research literature has contributed to understanding the mechanisms behind singing acoustics and perception. In addition to its complexity, the human voice is also the most popular musical instrument, as everybody sings and listens to vocal music. This has led to many practical engineering applications based on computational tasks such as singing assessment (linked to pitch, timbre and timing description), voice separation, singing synthesis, singer identification, singing style characterization, and query by humming/singing. Recent advancements in machine learning, in particular deep learning (DL), have provided a boost in performance in the mentioned computational tasks. However, these methods are still limited in terms of pitch resolution, separation quality, and naturalness of synthesized singing. In fact, the success of DL methods depends on three main factors: data, computing resources, and selected architectures.

- In terms of data, the performance of DL is directly connected to the amount and quality of available annotations. While data is more widely available in industrial settings, open datasets are still scarce due to privacy issues and industrial interest, the research community lacks of an agreed methodology for data gathering and annotation, and data needs are highly dependent on the computational task.

- Concerning computing, GPU availability is crucial for training complex architectures with high amounts of data. Computing resources are not always mentioned in scientific publications, and they provide some competitive advantages to researchers in large labs.
- Finally, in terms of DL architectures, there are still few studies on singing-specific algorithms that take advantage of knowledge on singing acoustics and perception.

Given the current research landscape, we need to address two crucial issues for our future research. First, we need to assess if the traditional link between singing processing and singing acoustics/perception research is lost because of the recent focus on deep learning. We believe that the efforts into explainable and optimized DL models might bring back this link in the future. Second, we need to consider if this boost in performance will have a different impact on listeners and singers, e. g. if singing synthesis becomes indistinguishable from human singing or if good quality singer impersonation is possible from polyphonic and noisy recordings. In this respect, our community has to be more aware of the social impact that our technologies might bring in the future.

3.10 Singing Information Processing

Masataka Goto (AIST – Tsukuba, JP)

License  Creative Commons BY 3.0 Unported license
© Masataka Goto

As music information research has continued to develop, research activities related to singing have become more vigorous. Such activities are attracting attention not only from a scientific point of view but also from the standpoint of commercial applications. Singing-related research is highly diverse, ranging from basic research on the features unique to singing to applied research such as that on singing synthesis, lyrics recognition, lyrics alignment, singer identification, retrieval of singing voices, singing skill evaluation, singing training, singing voice conversion, singing impression estimation, and the development of singer robots. I have named this broad range of singing-related studies “singing information processing” in 2008 (see [1] and [2]).

Since singing is one of the most important elements of music, singing information processing has a significant impact on society from the viewpoints of industry and culture. In fact, automatic pitch-correction technology for vocals is already used on a routine basis in the production of commercial music (popular music, in particular). It has become essential for correcting pitch at points in a song where the singer is less than skillful and for achieving a desired artificial effect. A function for evaluating (scoring) a person’s singing in the karaoke industry is also popular. More recently, singing-synthesis systems have become widely used, and people actively enjoy songs with synthesized singing voices as the main vocals.

During the Dagstuhl seminar, I gave an overview of this attractive research field. In particular, I covered the following examples for singing-related research:

- Singing synthesis including text-to-singing synthesis (VOCALOID), speech-to-singing synthesis (SingBySpeaking) and singing-to-singing synthesis (VocaListener).
- Robot singer (VocaListener + VocaWatcher).
- Singer identification.
- Retrieval of singing voices (VocalFinder).
- Creating hyperlinks between phrases in lyrics (Hyperlinking Lyrics).
- Lyrics alignment (LyricSynchronizer and Lyric Speaker).

I then discussed grand challenges such as ultimate singing analysis, superhuman singing synthesis, and perfect singing voice conversion. I finally addressed open questions to inspire an open discussion with participants from the viewpoint of technical issues and social impacts that singing technologies could give in the future.

References

- 1 Masataka Goto, Takeshi Saitou, Tomoyasu Nakano, and Hiromasa Fujihara. *Singing Information Processing Based on Singing Voice Modeling*. Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, USA, 2010, pp. 5506–5509.
- 2 Masataka Goto. *Singing Information Processing*. Proceedings of the 12th IEEE International Conference on Signal Processing (ICSP), Hangzhou, China, 2014, pp. 2431–2438.

3.11 Melody and Voice Processing – Some Thoughts on the Seminar

Frank Kurth (Fraunhofer FKIE – Wachtberg, DE)

License  Creative Commons BY 3.0 Unported license
© Frank Kurth

There is a multitude of real-life audio signals containing components with melody- or voice-like characteristics. With a research background from music audio, bioacoustics and speech processing, my interest has always been to explore synergies and cross-domain relationships in signal representation and analysis. Particularly for signal analysis, my interest in the seminar was to discuss how strong the impact of deep learning affects a research field like melody and voice processing, where generations of researchers have grown up with handcrafted and interpretable features (or, more generally, interpretable signal models). Relevant (and maybe provoking) questions were: Are interpretable signal features any longer necessary? How do learning-based approaches change the way we analyze signals and extract patterns? Are there subtasks in this field of MIR where machine learning already or in the future will outperform humans?

3.12 Try, Try, Try Again: Rehearsals as a Data Source

Cynthia Liem (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Cynthia Liem

(Comparative) analysis of musical performance and musical expression has, on the one hand, focused on commercially available recordings, and on the other hand on carefully conditioned experimental situations. In the first case, we usually only have an audio signal at our disposal which reflects the ultimate artistic intent of a performer and producer, although this intent is not always articulated explicitly. In the latter case, richer data is collected, but experimental conditions may not necessarily have created a ‘naturalistic’ environment. Whenever a musician is in progress of mastering a piece, multiple rehearsals are needed before the mastering is achieved. Within these rehearsals, multiple realizations of musical intent will be present; over time, they should ideally converge. As the musician progresses through the rehearsal, she also will likely go through various physical and mental states

of well-being, which may be evidenced in data that can be acquired during rehearsals. In summary, rehearsals may be a very interesting and data-rich source of information on musical expression, musician well-being, and developmental progress. At the Dagstuhl seminar, I discussed current data-acquisition and research activities at my lab that focus on this, both touching upon comparative performance analysis and physiological monitoring.

3.13 Singing Voice Separation: Recent Breakthroughs with Data-Driven Methods

Antoine Liutkus (Inria, University of Montpellier, FR)

License  Creative Commons BY 3.0 Unported license
© Antoine Liutkus

In my presentation, I gave an overview of different approaches that have been undertaken for the separation of vocals from music recordings in the past 40 years. First, I recalled that until recently, research on this topic largely focused on singing voice models based on physiological, acoustic, or musical aspects. In particular, researchers considered interpretable approaches that could be understood in terms of features of the singing voice and the accompaniment. Such model-based approaches, however, have never really met performance. Of course, it is always possible to find examples for which some given model is appropriate, but experience shows that any model proves inappropriate for the overwhelming majority of the other recordings.

Second, I explained that in this context, a new data-driven trend of research arose in the past five years in conjunction with deep learning. The singing voice is not described any more by some human-understandable model, but rather only through examples. In this setting, source separation systems are seen merely as mappings between mixtures and vocals. Training such systems has been made possible by the recent availability of dedicated datasets, where both isolated vocals and accompaniment music are available.

In the third part of my presentation, I demonstrated how the current state of the art impressively outperforms model-based approaches, based on the latest results from the international signal separation evaluation campaign (SiSEC). From these results followed a discussion about the current challenges on this topic, as well as how we should evaluate and compare contributions. Topics for discussion included the relative importance of models and their interpretability over trainable systems that lose interpretability but are much more effective at solving particular problems.

References

- 1 Estefania Cano, Derry Fitzgerald, Antoine Liutkus, Mark Plumbley, and Fabian-Robert Stöter. *Musical Source Separation: An Introduction*. IEEE Signal Processing Magazine, 36(1), 2019, pp. 31–40.
- 2 Zafar Rafii, Antoine Liutkus, Fabian-Robert Stöter, Stylianos Ioannis Mimilakis, Derry FitzGerald, and Bryan Pardo. *An Overview of Lead and Accompaniment Separation in Music*. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 26(8), 2018, pp. 1307–1335.
- 3 Fabian-Robert Stöter, Antoine Liutkus, and Nobutaka Ito. *The 2018 Signal Separation Evaluation Campaign*. International Conference on Latent Variable Analysis and Signal Separation, Springer, 2018, pp. 293–305.

3.14 Interactive Interfaces for Choir Rehearsal Scenarios

Meinard Müller (Universität Erlangen-Nürnberg, DE), Sebastian Rosenzweig (Universität Erlangen-Nürnberg, DE), and Frank Zalkow (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license

© Meinard Müller, Sebastian Rosenzweig, and Frank Zalkow

Joint work of Meinard Müller, Sebastian Rosenzweig, Frank Zalkow, Johannes Graulich (Carus-Verlag, DE)

Main reference Frank Zalkow, Sebastian Rosenzweig, Johannes Graulich, Lukas Dietz, El Mehdi Lemnaouar, Meinard Müller: “A Web-Based Interface for Score Following and Track Switching in Choral Music”. Demos and Late Breaking News of the International Conference on Music Information Retrieval (ISMIR), 2018

URL <https://www.audiolabs-erlangen.de/resources/MIR/2018-ISMIR-LBD-Carus>

Choral music is an essential part of our musical culture. Most choral singers practice their parts with traditional material, such as printed sheet music and CD recordings. Given recent advances in music information retrieval (MIR), important research questions are how and in which way new interfaces may enhance the rehearsal experience in particular for amateur choral singers. For example, score-following technology makes it possible to present audio and sheet music modalities synchronously. Furthermore, audio decomposition techniques may be useful to separate or enhance a specific voice (e. g., corresponding to the soprano part) that is relevant for a singer. Then, a choral singer’s rehearsal experience may be enhanced by switching between audio tracks of a multitrack recording. In collaboration with the Carus publishing house, a leading music publisher for religious and secular choral music with headquarters in Stuttgart, we explore the potential of such MIR technologies by building web-based prototypes for the interactive navigation and access of choral music recordings [4]. In particular, such interfaces should include personalization strategies that allow users to structure and analyze music recordings according to their specific needs, expectations, and requirements. Additionally, the integration of real-time feedback mechanisms concerning, e. g., rhythm, interaction, or intonation of the singers’ voices, may be useful components to make choir rehearsal preparations more effective [1].

In this context, a fundamental research topic is to investigate how and to what extent (partially) automated procedures can help to simplify the process of linking, decomposing, analyzing multimedia content. Even though there has been significant progress in MIR [2], the results of automatic alignment, voice separation, or music analysis procedures are still far from being perfect. As for commercial music applications, users are very critical regarding the quality of the presented music content. As for sheet music, for example, users often expect a visually appealing layout, where even small inaccuracies or distortions in the appearance of musical symbols may not be tolerable. Therefore, using software for automatically rendering sheet music such as Verovio [3] is often problematic when layout issues are of high importance. Similarly, when playing back a music recording along with showing a synchronized musical score, already small temporal asynchronies between the audio position and corresponding sheet music symbols may confuse the listener. Furthermore, feedback mechanisms on performance and intonation should work with high accuracy to satisfy a user’s expectation. Through close collaboration with the Carus publishing house, we explore the benefits and limitations of current MIR technologies in practical applications. This collaboration also offers numerous cross-connections to fields such as music education and musicology, which stimulates further interdisciplinary cooperations.

References

- 1 Helena Cuesta, Emilia Gómez, Agusín Martorell, and Felipe Loáiciga. *Analysis of Intonation in Unison Choir Singing*. Proceedings of the International Conference on Music Perception and Cognition (ICMPC), Graz, Austria, 2018.

- 2 Meinard Müller. *Fundamentals of Music Processing*. Springer Verlag, 2015.
- 3 Laurent Pugin, Rodolfo Zitellini, and Perry Roland. *Verovio: A library for Engraving MEI Music Notation into SVG*. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan, 2014, pp. 107–112.
- 4 Frank Zalkow, Sebastian Rosenzweig, Johannes Graulich, Lukas Dietz, El Mehdi Lemnaouar, and Meinard Müller. *A Web-Based Interface for Score Following and Track Switching in Choral Music*. Demos and Late Breaking News of the International Conference on Music Information Retrieval (ISMIR), 2018.

3.15 Singing Interfaces and Visualizations

Tomoyasu Nakano (AIST – Tsukuba, JP)

License  Creative Commons BY 3.0 Unported license
 © Tomoyasu Nakano

Technologies to automatically understand music and singing voices make it possible to develop systems enriching activities based on music. In such systems, interaction and visualization techniques play important roles. In this seminar, I introduced studies on interface development and information visualization based on signal processing and machine learning technologies. First, I discussed strategies to overcome errors introduced by automatic recognition approaches. One such strategy is based on human–computer interaction. For example, VocaListener, which imitates human singing expressions such as the fundamental frequency [1], integrates an interaction mechanism that lets a user easily correct lyrics-to-singing synchronization errors just by pointing them out. Furthermore, I discussed approaches for the effective visualization of music and singing. As an example of such research, I introduced TextTimeline which simultaneously visualizes words (lyrics) and acoustic features (intensity) [2]. By visualizing the time axis of the sound in a direction orthogonal to characters (vertical direction for horizontal text), TextTimeline can visualize them without changing the characters’ display position and without stretching the time axis of the sound.

References

- 1 Tomoyasu Nakano and Masataka Goto. *VocaListener: A Singing-to-Singing Synthesis System Based on Iterative Parameter Estimation*. Proceedings of the 6th Sound and Music Computing Conference (SMC), Porto, Portugal, 2009, pp. 343–348.
- 2 Tomoyasu Nakano, Jun Kato, and Masataka Goto. *TextTimeline: Visualizing Acoustic Features and Vocalized Timing along Display Text*. Proceedings of the 11th IEEE Pacific Visualization Symposium (IEEE PacificVis), Kobe, Japan, 2018.

3.16 Social Voice Processing

Juhan Nam (KAIST – Daejeon, KR)

License  Creative Commons BY 3.0 Unported license
 © Juhan Nam

Singing is a healthy activity that promotes not only physical conditions through exercise but also social or mental status by connecting people as an auditory medium. However, unless people are confident of their singing skills, they are reluctant to sing. Also, even

if the singing skill is good, recorded voices are often not satisfactory, compared against professionally processed voices. How can we encourage people to sing with more confidence, fun, and satisfaction?

We propose “social voice processing” as a concept of voice signal processing to transform input voice utilizing different renditions of voices for a given song in terms of singing skills and recording quality. Everyone has a different voice and a different level of singing skills. Therefore, every rendition of singing for a given piece of music is unique in terms of timbre, tempo, pitch, and dynamics. Even for the same singer, repeated performances have small differences. By leveraging karaoke apps or vast amount of singing voice recordings on online music content platforms, we can exploit such different singing performances for a given song to transform one’s voice as a digital audio effect.

One direction is mixing the multiple renditions. A traditional example is voice doubling as a recording technique. Mixing different voices can emulate a choir effect [1]. This can not only enrich the timbre but also suppresses singers’ concern about accurate pitch by diluting a prominent voice. Another direction is modifying the voice directly. A recent attempt is singing expression transfer that allows for exchanging musical expressions such as timing, pitch contours or energy between two voices [2]. This can improve one’s voice using skilled singers’ voices or make it worse for fun using intentionally poor voices. The two voices can be paired between multiple renditions from oneself, between friends, or between fans and professional singers [3]. There will be more possibilities by combining multiple singing voices and transferring different combinations of expressions among them. Furthermore, singing voice mixed with background music (e. g. commercial pop music) can be used along with melody extraction or source separation algorithms.

References

- 1 Mark Barry Dolson. *A Tracking Phase Vocoder and its Use in the Analysis of Ensemble Sounds*. Ph.D. thesis, California Institute of Technology, 1983.
- 2 Sangeon Yong and Juhan Nam. *Singing Expression Transfer from One Voice to Another for a Given Song*. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, 2018, pp. 151–155.
- 3 Masahiro Hamasaki, Masataka Goto, and Tomoyasu Nakano. *Songrium: A Music Browsing Assistance Service with Interactive Visualization and Exploration of A Web of Music*. Proceedings of the International Conference on World Wide Web (WWW), Seoul, Korea, 2014, pp. 523–528.

3.17 Automatic Singing Transcription for Music Audio Signals

Ryo Nishikimi (Kyoto University, JP)

License © Creative Commons BY 3.0 Unported license
© Ryo Nishikimi

Automatic singing transcription (AST) refers to estimating musical notes of a sung melody. Since the melody is the most salient part of music, the transcribed notes are useful for many MIR tasks such as singing voice generation, score-informed source separation, query-by-humming, and musical grammar analysis.

Since musical notes are represented as semitone-level pitches, onset score times, and note values, a singing voice should be quantized in the frequency and temporal directions. Most conventional methods aim to estimate a piano-roll representation by quantizing only pitches,

and note-value quantization (a.k.a rhythm transcription) has been studied independently. To integrate these tasks, it is necessary to associate frame-level spectrograms with tatum-level note values. To solve the AST problem, it is also necessary to accurately separate singing voice or directly estimate musical notes from music audio signals.

In this seminar, I introduced a statistical method of AST using hidden Markov models and demonstrated the transcriptions of recordings taken during the seminar. I plan to extend the model by integrating pitch and note-value quantizations and singing voice separation.

3.18 Dominant Melody Estimation and Singing Voice Separation

Geoffroy Peeters (Telecom ParisTech, FR)

License  Creative Commons BY 3.0 Unported license
© Geoffroy Peeters

In my presentation, I discussed recent research on dominant melody estimation and singing voice separation. Both topics have in common the use of Convolutional Neural Networks and skip-connections (U-Net). Furthermore, I reviewed possible input representations such as waveform, STFT, and HCQT representations [2, 3] as well as source/filter models [1]. Furthermore, I discussed how to automatically create large datasets annotated with pitch and lyrics using a student-teacher paradigm [4].

References

- 1 Dogac Basaran, Slim Essid, and Geoffroy Peeters. *Main Melody Extraction with Source-Filter NMF and C-RNN*. Proceedings of the International Society for Music Information Retrieval, Paris, France, 2018, pp. 82–89.
- 2 Alice Cohen-Hadria, Axel Roebel, and Geoffroy Peeters. *Improving Singing Voice Separation Using Deep U-Net and Wave-U-Net with Data Augmentation*. submitted to the IEEE International Conference on Acoustics, Speech, and Signal Processing, Brighton, UK, 2019.
- 3 Guillaume Doras, Philippe Esling, and Geoffroy Peeters. *On the Use of U-Net for Dominant Melody Estimation in Polyphonic Music*. Proceedings of the International Workshop on Multilayer Music Representation and Processing, Milan, Italy, 2019.
- 4 Gabriel Meseguer Brocal, Alice Cohen-Hadria, and Geoffroy Peeters. *Dali: A Large Dataset of Synchronized Audio, Lyrics and Pitch, Automatically Created Using Teacher-Student*. Proceedings of the International Society for Music Information Retrieval, Paris, France, 2018, pp. 431–437.

3.19 Russian/Ukrainian Traditional Polyphony: Musicological Research Questions and Challenges for MIR

Polina Proutskova (Queen Mary University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Polina Proutskova

As a performer and a collector of Russian/Ukrainian traditional polyphonic music, I have a practical application in mind which I would love to have and for which I would like to draw on the collective mind of the honored Dagstuhl participants.

These traditions are generally polyphonic but the relationship between the parts, particularly in older genres, is not harmonic but heterophonic: while in European traditions

the parts build consonant chords (are coordinated vertically), in Russian heterophony the voices are mainly constrained by the mode and are improvised in this mode, they are not coordinated vertically but conceived as “melodies” in their own right.⁶

Unfortunately, in a recording of more than three singers, it is difficult to make out all the parts by ear. Ethnomusicologists have used multi-channel recordings (where one voice is dominant) to capture the parts. See the website of the Polyphony Project⁷ for an example (scroll down for the multi-channel player). Yet such multi-channel recordings have been produced for only a small subset of the repertoire.

The inventiveness of folk singers is often breathtaking, and different ensembles have their unique styles. Can we model an ensemble’s style—the melodic lines created by each singer—based on multi-channel recordings of that ensemble? The goal would be to re-construct multi-channel recordings based on a recording of the whole ensemble. It would help me and my ensemble to learn songs based on archival recordings, for which no multi-channel takes were made.

In order to achieve that, our models will have to automatically transcribe the dominant voice in a messy choral recording, to learn the improvisational habits of a given singer, and to fit possible melodic lines for each of the singers so that together they build a construct closest to the recording of the whole ensemble. Also, our models would be able to produce a new voicing each time, true to the improvisational nature of the tradition.

If such a tool could be built, my ensembles IZBA (London) and Polynushka (Berlin) will be its grateful users and will help turn its output into real Russian folk songs. Conversely, the tool would help reconstruct a tradition nearly extinct, and bring archival recordings to life. And it might become a tool for more than reconstructing an old tradition, but for establishing a new one based on the principles of the old.

3.20 Aspects of Melodic Similarity

Preeti Rao (Indian Institute of Technology Bombay, IN)

License © Creative Commons BY 3.0 Unported license
© Preeti Rao

Joint work of Preeti Rao, Kaustuv Kanti Ganguli

The computation of “melodic similarity” remains central to MIR applications. We consider audio-based similarity involving the typical pipeline of deriving a suitable melodic representation from an audio “query” signal which is then compared with the reference representation to obtain a computed similarity concerning the reference. While acknowledging the importance of somehow modeling human similarity judgments, computational methods typically involve one of exact match, approximate match (i. e., exact matching applied to reduced forms) or categorical matching [1, 2]. Challenges also arise in the evaluation of melodic similarity measures. A compelling case is that of music that is naturally categorized based on a geographical basis such as folk tunes, or on a musicological basis such as the music of the raga and makam traditions. The similarity of pieces drawn from these genres depends on the local match of melodic motifs where the frequency of appearance of recognized motifs influences global similarity ratings [3]. Thus modeling of categorical matching at the time

⁶ Example: <https://www.youtube.com/watch?v=pP9p1p63BGE>

⁷ https://www.polyphonyproject.com/uk/song/BMI_UK18060360

scale of phrases or short melodic motifs appears relevant. In our work on raga music, we consider melodic motifs represented by continuous pitch curves. While trained musicians easily label the motifs in terms of solfege notation, it is all too common to find that the symbolic sequence representations are inadequate in discriminating ragas where identical sequences occur. With the raga considered to be a class of melodies described by tonal hierarchy and characteristic phrases, we find that the melodic motifs exhibit both intra-class and inter-class variations. A clustering study on a dataset of vocal concert recordings across two ragas with identical motifs in terms of solfege sequences reveals specific discriminating acoustic cues between the motifs. The cues derived from the dataset of pre-segmented phrases are consistent with musicological knowledge about the raga difference and, more importantly, serve to quantify the difference and make it exploitable for MIR. Perception experiments with trained musicians confirm the perceptual importance of the discovered cues [4, 5, 6]. The presented work points to fundamental questions around the segmentation of motifs, similarity measures between the melodic shapes (should this be some form of pitch distance or some higher-level cue-based comparison?), the process of learning cues from data, and finally applications to MIR and music education.

References

- 1 Emilios Cambouropoulos, Tim Crawford, and Costas S. Iliopoulos. *Pattern Processing in Melodic Sequences: Challenges, Caveats and Prospects*. Computers and the Humanities 35, 2001, pp. 9-21.
- 2 Alan Marsden. *Interrogating Melodic Similarity: A Definitive Phenomenon or the Product of Interpretation?* Journal of New Music Research, 41(4), 2012, pp. 323-335.
- 3 Anja Volk and Peter Van Kranenburg. *Melodic Similarity Among Folk Songs: An Annotation Study on Similarity-Based Categorization in Music*. Musicae Scientiae, 16(3), 2012, pp. 317-339.
- 4 Kaustuv Kanti Ganguli and Preeti Rao. *Towards Computational Modeling of the Ungrammatical in a Raga Performance*. Proceedings of the 18th International Society for Music Information Retrieval (ISMIR), 2017, pp. 39-45.
- 5 Kaustuv Kanti Ganguli and Preeti Rao. *Exploring Melodic Similarity in Hindustani Classical Music Through the Synthetic Manipulation of Raga Phrases*. Workshop on Cognitive Music Information Retrieval (CogMIR), New York, 2016.
- 6 Kaustuv Kanti Ganguli and Preeti Rao. *Imitate or Recall: How Do Musicians Perform Raga Phrases?* Proceedings of Frontiers of Research on Speech and Music (FRSM), Rourkela, India, 2017.

3.21 Extraction Techniques for Harmonic and Melodic Interval Analysis of Georgian Vocal Music

Sebastian Rosenzweig (Universität Erlangen-Nürnberg, DE), Meinard Müller (Universität Erlangen-Nürnberg, DE), and Frank Scherbaum (Universität Potsdam – Golm, DE)

License © Creative Commons BY 3.0 Unported license
© Sebastian Rosenzweig, Meinard Müller, and Frank Scherbaum

Polyphonic singing plays a vital role in many musical cultures. One of the oldest forms of polyphonic singing can be found in Georgia, a country located in the Caucasus region of Eurasia. The traditional three-voice chants are acknowledged as Intangible Cultural Heritage by the UNESCO. Being an orally transmitted culture, most of the sources are available as field recordings, which often are of rather poor audio quality. Musicologists typically research

on Georgian vocal music on the basis of manually created transcriptions of the recorded material. Such approaches are problematic since important tonal cues, as well as performance aspects, are likely to get lost in the transcription process. Within an interdisciplinary project, we apply and develop computational methods to support musicological research on traditional Georgian vocal music.

In this context, the non-tempered nature of the Georgian tuning is of particular interest and subject to many controversial discussions among musicologists [1, 2]. We try to contribute to this discussion by measuring melodic (horizontal) and harmonic (vertical) intervals of field recordings. From these two types of intervals, one then may obtain cues on the tonal organization of Georgian vocal music, as indicated in [3]. In our approach, we compute these intervals in three steps. First, we estimate the fundamental frequency (F0) trajectories for all three voices from the audio recordings. Second, we clean the extracted F0 trajectories using filtering techniques. Third, we determine harmonic and melodic intervals from suitable parts of the trajectories. This overall approach faces several challenges due to the polyphonic nature of the audio material. We report on several experiments to illustrate, handle, and circumvent some of these challenges. First, we show how one can improve the F0 estimation step by using informed and semi-automated (with user feedback) approaches [4]. Furthermore, we show how the F0 estimation problem can be significantly simplified when using headset and throat microphones additionally to conventional microphone types. We also discuss various filtering approaches to remove unstable and unreliable parts of the extracted F0 trajectories, which often correspond to pitch slides at the beginning and end of sung notes. Finally, we discuss the effect of the various F0 extraction and processing strategies on the derived harmonic and melodic intervals. In particular, we look at various interval distributions, which may serve as a basis for further musicological studies.

Besides contributing to the “Georgian scale controversy” our goal is to gain a deeper understanding of the interdependence of different sensor types, F0 extraction methods, and filtering techniques and their influence on the computed interval statistics.

References

- 1 Malkhaz Erkvanidze. *The Georgian Musical System*. Proceedings of the 6th International Workshop on Folk Music Analysis, Dublin, Ireland, 2016, pp. 74–79.
- 2 Zaal Tsereteli and Levan Veshapidze. *On the Georgian Traditional Scale*. Proceedings of the first International Symposium Traditional Polyphony, Tbilisi, Georgia, 2014, pp. 288–295.
- 3 Frank Scherbaum. *On the Benefit of Larynx-Microphone Field Recordings for the Documentation and Analysis of Polyphonic Vocal Music*. Proceedings of the 6th International Workshop on Folk Music Analysis, Dublin, Ireland, 2016, pp. 80–87.
- 4 Meinard Müller, Sebastian Rosenzweig, Jonathan Driedger, and Frank Scherbaum. *Interactive Fundamental Frequency Estimation with Applications to Ethnomusicological Research*. Proceedings of the AES International Conference on Semantic Audio, Erlangen, Germany, 2017.

3.22 Some Uncomfortable Statements about Melody Extraction

Justin Salamon (Adobe Research, US)

License  Creative Commons BY 3.0 Unported license
© Justin Salamon

In my Dagstuhl presentation, I raised two fundamental questions through the lens of melody extraction, even though these issues can be considered in the broader context of MIR research. Some of the statements in my talk might have annoyed you ... some might be wrong ... some might be true! What do *you* think?

■ Question: Is melody extraction “done right”?

Melody extraction has been an active topic of research in MIR for decades now, and yet there is still no consensus as to what a melody is. To illustrate this, I reviewed various definitions of melody found in musicology and music history, as well as definitions proposed by members of the MIR community, based on [1]. This review led to my first uncomfortable statement: *In MIR, a melody is whatever the annotations contain in the dataset I am using for my research.* That is, as a community, we often resort to “definition-by-annotation” to circumvent the challenge of estimating a musical concept that is inherently ambiguous. Should we change the way we think of, and evaluate melody extraction?

Next, I gave an overview of the existing datasets and evaluation metrics for melody extraction, their limitations, and the various efforts that have been made over the past few years to address these limitations. Importantly, I argued that the community has by-and-large ignored these efforts, and continues to use outdated datasets and evaluation metrics for melody extraction research. This led to my second uncomfortable statement: *Existing datasets for melody extraction are still (mostly) too small/artificial/homogenous, and (most) metrics in use have severe limitations ..., but we use them anyway!*

■ Question: What’s the point of melody extraction anyway?

In the second part of the talk, I presented some of the concepts and methodologies coming from the “lean startup” movement and contrasted them with equivalent processes in MIR research. In particular, I highlighted how the lean-startup methodology begins with a thorough customer discovery stage. This stage aims at identifying real problems, shared by a significant number of people, that require solutions. MIR research, in contrast, is sometimes driven by problems that may be interesting and challenging and may have “potential” applications, but that in practice have seen little applications outside of research. This led to my third uncomfortable statement: *There is a disconnect between MIR research (on melody extraction) and potential users of MIR technologies.*

References

- 1 Justin Salamon. *Melody Extraction from Polyphonic Music Signals*. Ph.D. Thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2013.

3.23 Computational Analysis of Traditional Georgian Vocal Music

Frank Scherbaum (Universität Potsdam – Golm, DE), Nana Mzhavanadze (Universität Potsdam – Golm, DE), Sebastian Rosenzweig (Universität Erlangen-Nürnberg, DE), Daniel Vollmer (Universität Potsdam – Golm, DE), Vlora Arifi-Müller (Universität Erlangen-Nürnberg, DE), and Meinard Müller (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
 © Frank Scherbaum, Nana Mzhavanadze, Sebastian Rosenzweig, Daniel Vollmer, Vlora Arifi-Müller, and Meinard Müller

Traditional multipart-singing is an essential component of the national identity of Georgia. It has been an active field of ethnomusicological research since more than 100 years, with a whole series of thematically very diverse research questions. In our contribution, we consider a computational approach, where we focus on the use of new and partially unconventional recording and analysis techniques to document and analyze this type of music. To circumvent the source separation problem for multiple singing voices in a natural singing environment, we explored the potential of recordings of skin vibrations close to the larynx (using larynx microphones) and the mastoid (using NAM microphones). In combination with conventional audio recordings (e. g., using high-quality headset microphones), these vibrational signals turned out very useful for subsequent computational analysis regarding a multitude of aspects including pitch tracking, tuning analysis, analysis of voice interaction, as well as for documentation and archiving purposes [1, 2, 3, 4, 5].

In rare cases, such as the Tbilisi State Conservatory Recordings of master chanter Artem Erkomaishvili in 1966 (which were recorded in an overdubbing mode), historical recordings can also be used to separate the individual voices using signal processing techniques and to investigate questions of historical performance practice, scales, and tuning, using computational techniques from the field of music information retrieval [6, 7]. Due to the existence of conventional transcriptions into Western staff notation by Shugliashvili [8], it became possible to compare different approaches to digitally represent this unique set of recordings statically (as images) as well as dynamically (as movies). One of the most exciting findings in the context of this work was the detection of numerous instances of melodic tuning adjustments (intonation changes to achieve particular harmonic intervals), in particular for harmonic fifths, ninths, octaves, and unison, the systematic investigation of which is now a topic of ongoing work.

References

- 1 Frank Scherbaum, Wolfgang Loos, Frank Kane, and Daniel Vollmer. *Body Vibrations as Source of Information for the Analysis of Polyphonic Vocal Music*. Proceedings of the International Workshop on Folk Music Analysis (FMA), Paris, France, 2015, pp. 89–93.
- 2 Scherbaum Frank. *On the Benefit of Larynx-Microphone Field Recordings for the Documentation and Analysis of Polyphonic Vocal Music*. Proceedings of the International Workshop on Folk Music Analysis (FMA), Dublin, Ireland, 2018, pp. 80–87.
- 3 Frank Scherbaum and Nana Mzhavanadze. *A New Archive of Multichannel-Multimedia Field Recordings of Traditional Georgian Singing, Praying, and Lamenting with Special Emphasis on Svaneti*. LaZAR-Database. <https://lazardb.gbv.de>
- 4 Frank Scherbaum, Nana Mzhavanadze, and Elguja Dadunashvili. *A Web-Based, Long-Term Archive of Audio, Video, and Larynx-Microphone Field Recordings of Traditional Georgian Singing, Praying and Lamenting with Special Emphasis on Svaneti*. Proceedings of the International Symposium on Traditional Polyphony, 2018.

- 5 Frank Scherbaum, Sebastian Rosenzweig, Meinard Müller, Daniel Vollmer, and Nana Mzhavanadze. *Throat Microphones for Vocal Music Analysis*. Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR), Paris, France, 2018.
- 6 Meinard Müller, Sebastian Rosenzweig, Jonathan Driedger, and Frank Scherbaum. *Interactive Fundamental Frequency Estimation with Applications to Ethnomusicological Research*. Proceedings of the AES Conference on Semantic Audio, Erlangen, Germany, 2017. <https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Erkomaishvili>
- 7 Frank Scherbaum, Meinard Müller, and Sebastian Rosenzweig. *Analysis of the Tbilisi State Conservatory Recordings of Artem Erkomaishvili in 1966*. Proceedings of the International Workshop on Folk Music Analysis (FMA), Málaga, Spain, 2017, pp. 29–36.
- 8 David Shugliashvili. *Georgian Church Hymns, Shemokmedi School*. Georgian Chanting Foundation & Tbilisi State Conservatory, 2014.

3.24 Measuring Intonation via Dissonance in Polyphonic Choral Singing

Sebastian J. Schlecht (Universität Erlangen-Nürnberg, DE), Christof Weiß (Universität Erlangen-Nürnberg, DE), Sebastian Rosenzweig (Universität Erlangen-Nürnberg, DE), and Meinard Müller (Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 3.0 Unported license
© Sebastian J. Schlecht, Christof Weiß, Sebastian Rosenzweig, and Meinard Müller

One of the central challenges in the performance of polyphonic choral singing is to sing in tune. Excellent intonation (i. e., a singer’s proper realization of pitch accuracy) requires many years to master. Especially in the challenging context of polyphonic singing, we see potential in supporting this learning task with computational techniques. Intonation may be defined either relative to an absolute pitch scale (such as equal temperament) or relative to other active harmonic sounds. Our current working hypothesis is that good intonation is achieved when the overall perceived tonal dissonance is minimized. We tested a dissonance measure developed by Sethares [1] based on the psychometric curves by Plomp and Levelt [2]. Sethares’ dissonance measure is computed from the relative frequency distance of all active partials and can be directly retrieved from audio recordings with partial tracking techniques [3]. Furthermore, we experimented with a dissonance- or tuning-measure based on the 12-tone equal-tempered scale. Inspired by [4], we accumulate the overall deviation of partial frequencies from an idealized 12-tone grid followed by a suitable normalization. As a test scenario, we compiled a small but diverse dataset of Anton Bruckner’s Gradual “Locus iste” in different performances, see [5] and the contribution described in Section 3.30.

At the Dagstuhl seminar, we raised questions about the potential of this approach in comparison to alternative methods. Further open questions included the separation of the composer’s intended dissonance from unintended dissonance introduced by the performance, the preferred quality of intonation in high dissonance harmonic contexts, and the role of dissonance in ensemble performances.

References

- 1 William A. Sethares. *Local Consonance and the Relationship Between Timbre and Scale*. Journal of the Acoustical Society of America, 94(3), 1993, pp. 1218–1228.
- 2 Rosina Plomp and Willem J. M. Levelt. *Tonal Consonance and Critical Bandwidth*. Journal of the Acoustical Society of America, 38(4), 1965, pp. 548–560.

- 3 Julian Neri and Philippe Depalle. *Fast Partial Tracking of Audio with Real-Time Capability Through Linear Programming*. Proceedings of the International Conference on Digital Audio Effects (DAFx), Aveiro, Portugal, 2018, pp. 326–333.
- 4 Tomoyasu Nakano, Masataka Goto, and Yuzuru Hiraga. *An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features*. Proceedings of the International Conference on Spoken Language Processing (Interspeech), 2006, pp. 1706–1709.
- 5 Helena Cuesta, Emilia Gómez, Agusín Martorell, and Felipe Loáiciga. *Analysis of Intonation in Unison Choir Singing*. Proceedings of the International Conference on Music Perception and Cognition (ICMPC), 2018.

3.25 Signal Processing for Multiple Fundamental Frequency Estimation

Li Su (Academia Sinica – Taipei, TW)

License  Creative Commons BY 3.0 Unported license
© Li Su

There are various competing methods for multiple fundamental frequency estimation (MF0E) of polyphonic recordings. Saying they are ‘competing’ does not imply that some of them will survive while others become useless in the future. Actually, what the different MF0E approaches compete for is not merely the accuracy achieved, but also their generalization capability and their applicability in real-world scenarios. By considering all these aspects, one also gets a better understanding of the benefits and limitations of recent machine learning methods (in particular, deep learning) and more traditional hand-crafted approaches in solving the MF0E problem.

In brief, machine learning and data-driven approaches for MF0E are driven by large-scale and well-annotated datasets. Hand-crafted or rule-based approaches, on the other hand, are driven by music theories and domain knowledge. Even though hand-crafted approaches might not be as competitive as the ones based on machine learning when evaluated against standard MIR datasets, traditional approaches may be still more useful in the wild, especially for case-dependent usage such as transcribing a class of non-Western music where training data is rare or even unavailable.

Both data-driven and rule-based approaches require techniques from signal processing. While rule-based MF0E usually employ signal-processing techniques to simplify the problem, data-driven methods need signal processing for generating suitable input (feature) representations as well for post-processing. For example, with a pure machine-learning framework for MF0E (framed as a classification problem), it may be hard to obtain F0 estimates at a high frequency resolution. Knowing the role of signal processing in MF0E, I discussed in my presentation the following two challenges.

- **Challenge: Instantaneous frequency (IF) estimation in general polyphonic music signals.** The definition of instantaneous frequency (IF) is itself an oxymoron. The Fourier paradigm assumes the stationarity of the analyzed signals, while the term instantaneous implies the signals are never stationary. In fact, in the sinusoidal model, the amplitude and frequency are not uniquely defined, if the stationary condition is not imposed. Analyzing multi-component signals is even more tricky because of the interference between overlapped components, which is unavoidable according to the Gabor-Heisenberg uncertainty principle. Voice and singing signals usually challenge the

basic assumptions of Fourier-based signal processing. For example, in choir singing, every note is actually an ensemble of sounds with nearby frequencies; it is neither a single tone nor several components with separable frequencies. How to break the quasi-stationary paradigm is a key to describe what the IF is in musical signal, and this topic is highly related to fundamental signal processing theories.

Another noteworthy advantage of signal processing is that smart use of signal processing can greatly reduce the data or computation resources in model training. A classic example is the pitch detection method combining frequency and periodicity representations, which was proposed by Peeters in [1]. It uses the temporal features to help suppress the unwanted harmonics of a component and enhances true F0 peaks. Such an approach is found not only useful for music signals but also for the MF0E tasks in biomedical signals [2, 3]. An extension of this approach is the recently proposed multi-layer cepstrum (MLC), which performs the Fourier transform, nonlinear scaling and a high-pass filter recursively to achieve iterative purification of F0 information [4]. Preliminary studies also show its potential in analyzing choir singing.

- **Challenge: Use of signal processing in transcribing non-Western music, taking Taiwanese aboriginal music ‘Pasibutbut’ as an example.** There are more than ten aboriginal tribes living in Taiwan. Studies on Taiwanese aboriginal music started in the mid-20th century. According to the studies, most of the tribes’ choir music is heterophonic, and some tribes such as Bunun and Amis even have the tradition of polyphonic choir singing. Pasibutbut (meaning “Praying for a Rich Harvest”) is a classic example of polyphonic singing of the Bunun tribe. The leading voice (named ‘Mahusngas’) gradually increases the pitch while the other three voices follow the leading voice singing lower than the leading voice by a minor third, perfect fourth, and perfect fifth, respectively. In the experiments, we found that the MLC method outputs the pitch contours of each part in a resolution better than humans can do. Processing the contours by an unsupervised clustering method such as DBSCAN results in note transcriptions without any labeled training data. In conclusion, a smart signal processing strategy can reduce efforts in labeling large amounts of data while giving satisfactory results for this type of non-Western music.

References

- 1 Geoffroy Peeters. *Music Pitch Representation by Periodicity Measures Based on Combined Temporal and Spectral Representations*. IEEE Proceedings of the International Conference on Acoustics Speech and Signal Processing Proceedings (ICASSP), Toulouse, France, 2006, pp. 53–56.
- 2 Li Su and Yi-Hsuan Yang. *Combining Spectral and Temporal Representations for Multipitch Estimation of Polyphonic Music*. IEEE/ACM Transactions Audio, Signal Language Processing (TASLP), 23(10), 2015, pp. 1600–1612.
- 3 Chen-Yun Lin, Li Su, and Hau-tieng Wu. *Wave-Shape Function Analysis – When Cepstrum Meets Time-Frequency Analysis*. Journal of Fourier Analysis and Applications, 24(2), 2018, pp. 451–505.
- 4 Chin-Yun Yu and Li Su. *Multi-layered Cepstrum for Instantaneous Frequency Estimation*. Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, USA, 2018.

3.26 Augmented Vocal Production towards New Singing Style Development

Tomoki Toda (Nagoya University, JP)

License  Creative Commons BY 3.0 Unported license
© Tomoki Toda

Singers can produce attractive singing voices by expressively controlling pitch, dynamics, rhythm, and voice timbre. However, individual singers have their own limitations to control these components widely owing to physical constraints in speech production. For instance, it is physically difficult to change their own voice timbre into that of another specific singer. Moreover, if they suffered from vocal disorder, they would be unable to produce singing voices. If singers could freely produce singing voices as they want beyond their physical constraints, it would open up entirely new ways to express a greater variety of expression.

Towards the development of techniques to augment our speech production mechanism, I have studied a real-time statistical voice conversion technique for more than 15 years. Statistical voice conversion is a technique based on machine learning to modify a speech waveform for converting non- or para-linguistic information while keeping linguistic information unchanged. Its real-time implementation has been successfully achieved by incorporating real-time signal processing. This technique makes it possible for us to produce speech and singing voices beyond our physical constraints, and therefore, it has great potential to develop various applications to break down the existing barriers in our speech production. I have named this technique “augmented speech production” [1], and have developed some applications for augmenting vocal production, such as vocal effector [2] and singing-aid for laryngectomees [3].

In the Dagstuhl seminar, I gave an overview of augmented speech production techniques, showed some applications in singing voice production including a demo system of vocal effector, and addressed open questions to inspire an open discussion with participants.

Acknowledgements: This work was partly supported by JST, PRESTO Grant Number JPMJPR1657.

References

- 1 Tomoki Toda. *Augmented Speech Production Based on Real-Time Statistical Voice Conversion*. Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, USA, 2014, pp. 755–759.
- 2 Kazuhiro Kobayashi, Tomoki Toda, and Satoshi Nakamura. *Intra-Gender Statistical Singing Voice Conversion with Direct Waveform Modification Using Log-Spectral Differential*. Speech Communication, Vol. 99, 2018, pp. 211–220.
- 3 Kazuho Morikawa and Tomoki Toda. *Electrolaryngeal Speech Modification Towards Singing Aid System for Laryngectomees*. Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, Malaysia, 2017, pp. 610–613.

3.27 Useless Evaluation versus User-less Evaluation

Julián Urbano (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Julián Urbano

Performance metrics are one of the keystones in our everyday research. They (should) serve us as abstractions of how our algorithms are used in real use cases, and the numbers we compute with them are the evidence we present to judge the merit of our research and decide what works and what does not. However, do they tell us anything useful? Most of our metrics are algorithm-oriented, and that makes sense. Many of our algorithms are just smaller parts in bigger systems and so it makes sense to treat them in a purely system-oriented fashion. However, very often we have real people at the other end, and how they perceive the output of our algorithms, or how useful it is for them, cannot be measured in a system-oriented fashion. For instance, probably not all kinds of mistakes are perceived the same, and two outputs with the same accuracy may be perceived entirely differently depending on how errors are arranged throughout the piece. How do real users perceive the output from our algorithms? Should we start devising new user-oriented measures? Can we do it without constant human intervention?

3.28 Computational Modeling of Melody

Anja Volk (Utrecht University, NL)

License  Creative Commons BY 3.0 Unported license
© Anja Volk

Joint work of Anja Volk, Marcelo Rodriguez-Lopez, Peter van Kranenburg, Iris Yuping Ren

In contrast to harmony and rhythm, melody is not considered a “basic musical structure” [11] in music theory. Accordingly, while there exist many theories on harmony and rhythm, theories on melodies are sparse. The computational modeling of melody in MIR and computational musicology has focused on topics such as modeling melodic similarity [11], melodic segmentation [10], the stability of melodic features over the course of oral transmission in folk music [3], modeling the role of melodies for listeners’ expectations [8], and discovering prototypical contours and patterns [2, 8].

Discovering repeated patterns which are musically meaningful is a specifically challenging task, as usually algorithms discover much more patterns than humans [5]. However, repeated patterns have been shown to be crucial for important aspects of melody such as similarity [13, 1] and segmentation [10], and can be used for discovering repeated sung phrases in audio corpora [4]. I consider it an interesting direction of research on how the computational modeling of melody, specifically the aspect of melodic patterns, in symbolically annotated music and in recordings can cross-pollinate each other in order to improve challenging tasks such as melodic contour extraction. Integrating audio and symbolic approaches to modelling melodic aspects would not only contribute to solving specific MIR tasks related to melody, but might contribute to a broader theoretization of the phenomenon of melody also in the context of musicology and cognition.

References

- 1 Peter Boot, Anja Volk, and W. Bas de Haas. *Evaluating the Role of Repeated Patterns in Folk Song Classification and Compression*. *Journal of New Music Research*, 45(3), 2016, pp. 223–238.
- 2 Kaustuv Ganguli and Preeti Rao *Discrimination of Melodic Patterns in Indian Classical Music*. National Conference on Communications, IIT Bombay, India 21 (1), 2015, pp. 1–6.
- 3 Berit Janssen, Peter van Kranenburg, and Anja Volk. *Finding Occurrences of Melodic Segments in Folk Songs Employing Symbolic Similarity Measures*. *Journal of New Music Research*, 46(2), 2017, pp. 118–134.
- 4 Nadine Kroher, Aggelos Pikrakis, and Jose-Miguel Díaz-Báñez, *Discovery of Repeated Melodic Phrases in Folk Singing Recordings*. *IEEE Transactions on Multimedia* 20(6), 2018, pp. 1291–1304.
- 5 Iris Yuping Ren, Hendrik Vincent Koops, Anja Volk, and Wuoter Swierstra. *In Search of the Consensus Among Musical Pattern Discovery Algorithms*. Proceedings of the 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017, pp. 671–678.
- 6 Iris Yuping Ren, Anja Volk, Wouter Swierstra, and Remco C. Veltkamp. *Analysis by Classification: A Comparative Study of Annotated and Algorithmically Extracted Patterns in Symbolic Music Data*. Proceedings of the 19th International Society for Music Information Retrieval Conference, Paris, France, 2018, pp. 539–546.
- 7 Iris Yuping Ren, Hendrik Vincent Koops, Dimitrios Bountouridis, Anja Volk, Wouter Swierstra, and Remco C. Veltkamp. *Feature Analysis of Repeated Patterns in Dutch Folk Songs using Principal Component Analysis*. Proceedings of the 8th International Workshop on Folk Music Analysis, Thessaloniki, Greece, 2018, pp. 86–87.
- 8 Marcelo E. Rodríguez-López and Anja Volk. *Symbolic Segmentation: A Corpus-Based Analysis of Melodic Phrases*. Proceedings of the 10th International Symposium on Computer Music Modeling and Retrieval (CMMR), Laboratoire de Mécanique et d’Acoustique, Marseille, France, 2013, pp. 548–557.
- 9 Marcelo E. Rodríguez-López and Anja Volk: *Melodic Segmentation Using the Jensen-Shannon Divergence*. Proceedings of the 11th International Conference on Machine Learning and Applications, Boca Raton, USA, 2012, pp. 351–356.
- 10 Marcelo E. Rodríguez López and Anja Volk. *Location Constraints for Repetition-Based Segmentation of Melodies*. Proceedings of the 5th International Conference in Mathematics and Computation in Music (MCM), Springer, 2015, pp. 73–84.
- 11 David Temperley, *The Cognition of Basic Musical Structures*. MIT Press, 2001.
- 12 Peter van Kranenburg, Anja Volk, Frans Wiering, and Remco C. Veltkamp. *Musical Models for Folk-Song Melody Alignment*. Proceedings of the International Society on Music Information Retrieval Conference, Kobe, Japan, pp. 507–512.
- 13 Anja Volk and Peter van Kranenburg. *Melodic Similarity Among Folk Songs: An Annotation Study on Similarity-Based Categorization in Music*. *Musicae Scientiae*, 16(3), 2012, pp. 317–339.

3.29 Singing Voice Modelling for Language Learning

Ye Wang (National University of Singapore, SG)

License  Creative Commons BY 3.0 Unported license
© Ye Wang

Singing is a popular form of entertainment, as evidenced by the millions of active users of Karaoke apps like Smule’s Sing! and Tencent’s Quanmin K Ge. Singing is presumed to be the oldest form of music making and can be found in human cultures around the world. However, singing can be more than just a source of entertainment: parents sing nursery rhymes to their young children to help them learn their first language, music therapists use singing to help aphasia patients speak again, and medical studies have revealed that singing, in general, has many health benefits. Consequently, computational methods for singing analysis have emerged as an active research topic in the music information retrieval community.

Pedagogical research has shown that actively singing in a foreign language helps with pronunciation, vocabulary acquisition, retention, fluency, and cultural appreciation [1]. Inspired by this scientific discovery, we have developed a novel multi-language karaoke application called SLIONS (Singing and Listening to Improve Our Natural Speaking), designed to foster engaging and joyful language learning [2]. We followed a user-centered design process that was informed by conducting interviews with domain experts and by conducting usability tests among students. The key feature of SLIONS is an automatic speech recognition (ASR) tool used for objective assessment of sung lyrics, which provides students with personalized, granular feedback based on their singing pronunciation.

During its proof of concept phase, SLIONS employed Google’s ASR technology to evaluate sung lyrics. However, this solution lacks technical depth and has several critical limitations. First, Google ASR is proprietary technology and is effectively a black box. As a result, it is impossible for us to understand precisely why it succeeds in evaluating certain sung lyrics but fails in others. This not only prevents us from gaining insights into the underlying models but also affects SLIONS’ value in real-world applications. It is also impossible for us to modify Google’s ASR technology even though we wish to use SLIONS for widely varying applications that range from language learning to melodic intonation therapy. Google ASR technology is designed for speech recognition and is suboptimal for analyzing singing voice, as the characteristics of sung utterances differ from those of spoken utterances. Therefore it is desirable to investigate better and more versatile computational methods for objective assessment of sung lyrics. Furthermore, it is also useful to address some important human-computer interaction (HCI) questions. For example, while previous studies have shown that singing can help with learning pronunciation, the critical question of which factors are essential for not only improving pronunciation but also maintaining engagement during singing exercises remains. It is vital to design interface/interaction features that support both learning and engagement aspects.

Although speech and singing share a common voice production organ and mechanism, singing differs from speech in terms of pitch variations, possible extended vowels, vibrato, and more. It is interesting to exploit the similarities between speech and singing in order to employ existing methods/tools and datasets in the relatively mature ASR field while also developing new methods to address the differences. To this end, we have created and published the NUS Sung and Spoken Lyrics Corpus, a small phonetically annotated dataset of voice utterances [3]. Furthermore, we have also attempted to address the problem of lyrics and singing alignment [4, 5, 6, 7], evaluation of sung lyrics [8, 9], and intelligibility of sung lyrics [10]. While many challenges remain as to adequately modeling and analyzing singing

voice for real-world applications such as language learning, our efforts are already pointing the way towards a robust, versatile model that can enable the automatic evaluation of sung utterance pronunciation.

References

- 1 Arla J. Good, Frank A. Russo, and Jennifer Sullivan. *The Efficacy of Singing in Foreign-Language Learning*. *Psychology of Music*, 43(5), 2015, pp. 627–640.
- 2 Dania Murad, Riwu Wang, Douglas Turnbull, and Ye Wang. *SLIONS: A Karaoke Application to Enhance Foreign Language Learning*. Proceedings of the ACM International Conference on Multimedia, Seoul, Korea, 2018, pp. 1679–1687.
- 3 Zhiyan Duan, Haotian Fang, Bo Li, Khe Chai Sim, and Ye Wang. *The NUS Sung and Spoken Lyrics Corpus: A Quantitative Comparison of Singing and Speech*. Proceedings of the IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Kaohsiung, Taiwan, 2013, pp. 1–9.
- 4 Chitrlekha Gupta, Rong Tong, Haizhou Li, and Ye Wang. *Semi-Supervised Lyrics and Solo-Singing Alignment*. Proceedings of International Society for Music Information Retrieval Conference (ISMIR), Paris, France, 2018, pp. 600–607.
- 5 Min-Yen Kan, Ye Wang, Denny Iskandar, Tin Lay Nwe, and Arun Shenoy. *LyricAlly: Automatic Synchronization of Textual Lyrics to Acoustic Music Signals*. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2), 2008, pp. 338–349.
- 6 Denny Iskandar, Ye Wang, Min-Yen Kan, and Haizhou Li. *Syllabic Level Automatic Synchronization of Music Signals and Text Lyrics*. Proceedings of the ACM International Conference on Multimedia, Santa Barbara, USA, 2006, pp. 659–662.
- 7 Ye Wang, Min-Yen Kan, Tin Lay Nwe, Arun Shenoy, and Jun Yin. *LyricAlly: Automatic Synchronization of Acoustic Musical Signals and Textual Lyrics*. Proceedings of the ACM International Conference on Multimedia, New York, USA, 2004, pp. 212–219.
- 8 Chitrlekha Gupta, Haizhou Li, and Ye Wang. *Automatic Pronunciation Evaluation of Singing*. Proceedings of the International Conference on Spoken Language Processing (Interspeech), Hyderabad, India, 2018, pp. 1507–1511.
- 9 Chitrlekha Gupta, David Grunberg, Preeti Rao, and Ye Wang. *Towards Automatic Mispronunciation Detection in Singing*. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Suzhou, China, 2017, 390–396.
- 10 Karim M. Ibrahim, David Grunberg, Kat Agres, Chitrlekha Gupta, and Ye Wang. *Intelligibility of Sung Lyrics: A Pilot Study*. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Suzhou, China, 2017, pp. 686–693.

3.30 Analyzing and Visualizing Intonation in Polyphonic Choral Singing

Christof Weiß (Universität Erlangen-Nürnberg, DE), Sebastian J. Schlecht (Universität Erlangen-Nürnberg, DE), Sebastian Rosenzweig (Universität Erlangen-Nürnberg, DE), and Meinard Müller (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Christof Weiß, Sebastian J. Schlecht, Sebastian Rosenzweig, and Meinard Müller

Unaccompanied vocal music constitutes the nucleus of Western art music and the starting point of polyphony’s evolution. Despite an increasing number of exciting studies [1, 2, 3], many facets of polyphonic a cappella singing are yet to be explored and understood. In some preliminary experiments, we made first attempts to investigate and visualize intonation aspects in choral singing. At this moment, many questions arise that are highly interrelated:

Which effects occur at different temporal levels (global reference tuning, pitch drifts, local interval errors)? What are the influences of singing conditions (solo singers, multiple singers per part, feedback and interaction), level of training (professional singers vs. amateurs), and acoustic conditions (room size, reverb)? What is the role of the musical composition and how does it affect tuning adjustment (harmony, complexity of chords, interactions of parts, and the influence of overtones)? Finally, how are these effects perceptually relevant to listeners?

In order to systematically study such questions, we compiled a small but diverse dataset of Anton Bruckner’s Gradual “Locus iste” (WAB 23) in different performances. Our examples comprise a 16-singer multi-track recording from the Choral Singing Dataset [3] as well as several commercial and non-commercial performances. Furthermore, we generated sine-tone renditions of the piece with different pitch accuracy using random pitch deviations. We experimented with several types of visualization for investigating pitch trajectories in relation to the score in an intuitive way. Furthermore, we tested more general measures of consonance and tuning quality, which do not require score information (see also Section 3.24). Such visualizations may have high potential as supporting tools for choir training and rehearsals.

References

- 1 Jiajie Dai and Simon Dixon. *Analysis of Interactive Intonation in Unaccompanied SATB Ensembles*. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Suzhou, China, 2017, pp. 599–605.
- 2 Matthias Mauch, Klaus Frieler, and Simon Dixon. *Intonation in Unaccompanied Singing: Accuracy, Drift and a Model of Reference Pitch Memory*. *Journal of the Acoustical Society of America*, 136(1), 2014, pp. 401–411.
- 3 Helena Cuesta, Emilia Gómez, Agusín Martorell, and Felipe Loáiciga. *Analysis of Intonation in Unison Choir Singing*. Proceedings of the International Conference on Music Perception and Cognition (ICMPC), 2018.

3.31 Melody Extraction for Melody Generation

Yi-Hsuan Yang (Academia Sinica – Taipei, TW)

License  Creative Commons BY 3.0 Unported license
© Yi-Hsuan Yang

Recent years have witnessed a growing interest in applying machine learning algorithms to automatically compose or generate music in the symbolic domain [1, 4, 5, 6]. As melody is a core part of a music piece, many prior approaches have focused on generating a melody, considered as a sequence of notes (specified at a semitone resolution) [2]. Recently, this is usually done by collecting a large number of melodies in a symbolic format such as XML, finding a way to represent the melodies computationally, and then training a neural network to learn to generate melodies, either with conditions (e. g., chords [4]) or no conditions.

However, we note that there are two significant limitations of the approach above. First, not all the existing melodies in the world have been or can be digitally stored in a symbolic format, and are available to us. Relying on melodies in the symbolic format restricts the quantity as well as the diversity of the training dataset. Second, melodies in semitones lack performance-level attributes such as variations in pitch, dynamics, and note duration. Accordingly, to render realistic audio, one needs to either invite a human musician to play the melody, or to build another machine learning model to synthesize the audio from the given melody.

The PerformanceRNN [3] proposed by Google Magenta, and some follow-up research, attempts to address the second issue mentioned above by generating music with performance-level attributes in a single pass. However, this has been limited to piano music only, and not the melody for general music.

We propose here the idea of using the result of audio-domain melody extraction (e. g., [7]) to learn to compose/generate melodies. Such audio-domain melody extraction algorithms can be applied to any music piece as long as there are audio recordings, thereby bypassing the difficulty of obtaining symbolic data. Moreover, the target output of such algorithms is pitch specified in Hertz (rather than in semitones), comprising performance-level attributes in pitch and note duration. This makes it possible to learn to generate melodies of diverse styles and to learn to generate melodies with expressive qualities.

One can further apply note tracking algorithms (e. g., [8]) to convert the melody contour in Hertz to a melody note sequence in semitones, and use both the Hertz and semitone versions to train melody composition and generation models.

The PerformanceNet [9] is a convolutional neural network model we recently proposed to convert a piano-roll-like symbolic-domain musical score to the magnitude spectrogram, which can then be rendered to audio. Following similar ideas, one can build a model that generates the melody in semitones first, and then, based on that, to generate the melody in Hertz.

The idea can be more broadly considered as a “transcription first, and then generation” approach. It is a promising use case of audio-domain melody extraction algorithms, and it might lead to new progress in music generation research.

References

- 1 Bob L. Sturm, João Felipe Santos, Oded Ben-Tal, and Iryna Korshunova. *Music Transcription Modelling and Composition Using Deep Learning*. Proceedings of the Conference on Computer Simulation of Musical Creativity, Huddersfield, UK, 2016.
- 2 Elliot Waite. *Project Magenta: Generating Long-Term Structure in Songs and Stories*. 2016. <https://magenta.tensorflow.org/2016/07/15/lookback-rnn-attention-rnn>
- 3 Ian Simon and Sageev Oore. *Performance RNN: Generating Music with Expressive Timing and Dynamics*. 2017. <https://magenta.tensorflow.org/performance-rnn/>
- 4 Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang. *MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation*. Proceedings of the International Society for Music Information Retrieval Conference, Suzhou, China, 2017, pp. 324–331.
- 5 Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. *MuseGan: Multi-Track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment*. Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018, pp. 34–41.
- 6 Jean-Pierre Briot, Gaëtan Hadjeres, and François Pachet. *Deep Learning Techniques for Music Generation: A Survey*. arXiv preprint, 2017. <http://arxiv.org/abs/1709.01620>
- 7 Justin Salamon, Emilia Gómez, Daniel P. W. Ellis, Gaël Richard. *Melody Extraction from Polyphonic Music Signals: Approaches, Applications and Challenges*. IEEE Signal Processing Magazine, 31(2), 2014, pp. 118–134.
- 8 Yuan-Ping Chen, Ting-Wei Su, Li Su, and Yi-Hsuan Yang. *TENT: Technique-Embedded Note Tracking for Real-World Guitar Solo Recordings*. Transactions of the International Society for Music Information Retrieval, 2019.
- 9 Bryan Wang and Yi-Hsuan Yang. *PerformanceNet: Score-To-Audio Music Generation with Multi-Band Convolutional Residual Network*. Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, USA, 2019.

3.32 Finding Musical Themes in Western Classical Music Recordings

Frank Zalkow (Universität Erlangen-Nürnberg, DE), Stefan Balke, and Meinard Müller (Universität Erlangen-Nürnberg, DE)

License  Creative Commons BY 3.0 Unported license
© Frank Zalkow, Stefan Balke, and Meinard Müller

Many pieces from Western classical music contain short melodies or musical gestures that are especially prominent and memorable, so-called musical themes. Finding such themes in audio recordings is a challenging cross-modal retrieval scenario. In such a setting, the query is a symbolic encoding of the theme's melody, and the database is a collection of classical music recordings. In this scenario, we face several problems: the difference in modality, differences in tuning, transposition, and tempo, as well as the difference in polyphony between a query and database document [1].

Usually, one employs common mid-level representations for performing retrieval tasks with different modalities. In particular, one may use chroma features, which measure the energy in the twelve chromatic pitch class bands. The difficulty due to the difference in polyphony could be bypassed with accurate melody extraction methods [4]. Those methods often work on so-called salience representations, which are time–frequency representations with enhanced tonal frequency components [2, 3]. However, for polyphonic classical music, melody extraction often fails. Therefore, we propose not to extract the melodies, but to directly map the salience representations to chroma features before performing the retrieval [6]. As an alternative, we suggest learning common mid-level representations for both query and database with a data-driven approach, e. g., using deep learning with the triplet loss [5].

References

- 1 Stefan Balke, Vlora Arifi-Müller, Lukas Lamprecht, and Meinard Müller. *Retrieving Audio Recordings Using Musical Themes*. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 281–285.
- 2 Meinard Müller. *Fundamentals of Music Processing*. Springer Verlag, 2015.
- 3 Justin Salamon and Emilia Gómez. *Melody Extraction from Polyphonic Music Signals Using Pitch Contour Characteristics*. IEEE Transactions on Audio, Speech, and Language Processing, 20(6), pp. 1759–1770, 2012.
- 4 Justin Salamon, Emilia Gómez, Daniel P. W. Ellis, and Gaël Richard. *Melody extraction from polyphonic music signals: Approaches, applications, and challenges*. IEEE Signal Processing Magazine, 31(2), pp. 118–134, 2014.
- 5 Florian Schroff, Dmitry Kalenichenko, and James Philbin. *FaceNet: A Unified Embedding for Face Recognition and Clustering*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 815–823.
- 6 Frank Zalkow, Stefan Balke, and Meinard Müller. *Evaluating Salience Representations for Cross-Modal Retrieval of Western Classical Music Recordings*. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, United Kingdom, 2019.

4 Working Groups

4.1 Data Sets: What is Missing and How do We Get It?

Participants of Dagstuhl Seminar 19052

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

This working group considered the data needs of singing research, specifically the lack of isolated vocal data for singing voice synthesis and other singing voice research. We proposed that the first step in addressing the issue of data paucity is to establish what data we already have and where exactly the holes are. To this end, we discussed developing a shared document (perhaps on Google) that lists existing data. We envision this as a bootstrapped (and free) version of the Linguistic Data Consortium⁸. We also considered the various ways in which data could be collected (including recordings by researchers, collaborating with karaoke services and other relevant companies, and using source-separation on existing recordings), how we may deal with related copyright issues, and how much data we need for different tasks. This led us to consider how diverse our data needs to be in terms of pitch range, vocabulary, singer type, and genre as well as how it should be annotated with respect to musical and lyrical content. We concluded that, in addition to creating a resource for listing available data sets, we should also work on creating community-sourced guidelines on best practices on curating and annotating singing data sets.

4.2 Deep Learning Versus Acoustic Models

Participants of Dagstuhl Seminar 19052

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

In this working group, we discussed how different types of systems can represent the acoustical and physical properties of singing. The main focus was on the relationship between implicit knowledge, as encoded in a Deep Neural Network (DNN) trained on annotated data, and explicit knowledge, as built into hand-crafted signal processing approaches.

As the first central question, we discussed to which extent current Deep Learning (DL) approaches exploit acoustic models of singing production in the design of their model architecture. We noticed that there is only a small number of studies that explicitly exploit such knowledge. For example, in the work by Basaran et al. [1], a source-filter model is used as an input to a CNN. Another example is the WaveNet vocoder [2]. For the majority of works, knowledge is only given implicitly to the system via training on suitable data. In this context, data augmentation strategies play a crucial role to enforce that a system possesses certain invariance properties.

As the second major question, we wondered what kind of knowledge about singing production could potentially be derived from trained systems. For example, can we learn an acoustic or physical model of singing production from a DL system? We agreed that deriving any explicit knowledge from a trained model is generally hard. The architecture itself

⁸ <https://catalog.ldc.upenn.edu>

provides little insights; indeed many architectures are used successfully for many different problems within MIR. Further points that arose in our discussion touched the challenge of data collection. As one strategy, semi-supervised training procedures were proposed, for example, by exploiting information from synchronized videos or by making use of suitable embeddings.

References

- 1 Dogac Basaran, Slim Essid, and Geoffroy Peeters. *Main Melody Estimation with Source-Filter NMF and CRNN*. Proceedings of the International Society for Music Information Retrieval Conference, Paris, France, 2018, pp. 82–89.
- 2 Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. *Wavenet: A generative model for raw audio*. Electronic preprint (arXiv:1609.03499), 2016.

4.3 Insights

Participants of Dagstuhl Seminar 19052

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

In this working session, we addressed the question of what it means to provide “insights” into tasks, techniques, data, and any aspect related to our research area. We started by discussing how one may improve and evaluate the insights provided by ISMIR papers. We agreed on the usefulness of community guidelines that may help authors, reviewers, and meta-reviewers to improve and evaluate the usefulness of scientific contributions. The following list contains some of our suggestions:

- A paper should go beyond stating something like “we got x% accuracy on this dataset.”
- The aspect of “insights” could be added in author guidelines. Furthermore, it could be added in the ISMIR paper template.
- Good/bad examples of result and conclusion sections should be provided. Such examples could be collected via “crowdsourcing.”
- Guidelines could be provided on the ISMIR website as early as possible so that potential authors are ready and prepared.
- As for reviewing, the review form could be expanded by a specific question about the insights of a submission. Reviewers should explicitly comment on technical, conceptual and other insights provided by a submission.
- Meta-reviewers should evaluate a paper according to its insights. Such insights may refer to different aspects, e. g. machine learning insights or general MIR insights.
- The discussion could be continued within the broader ISMIR community, e. g., by means of a shared Google doc or an email thread.

In the second half of our group discussion, we had a controversial debate about the importance of sharing research code as part of publications. In particular, we discussed whether open source code is useful or necessary for improving the insight provided by MIR papers. A variety of opinions and issues were mentioned, which can be roughly summarized as follows:

- Code must be shared. Otherwise, the research is not reproducible, and thus provides little insight. Without sharing the code, how can the community be confident the insight provided by a paper is correct?

- Code sharing should be encouraged but not mandatory. Sharing the code is its own reward, as it increases impact. Code sharing increases reproducibility, which in turn increases impact.
- Code sharing is nice, but far less important than providing a good description of the method and analysis of the results. In other words, sharing code in lieu of a good description of the method is not acceptable. Papers should not be rejected for not sharing code.
- Code sharing is irrelevant; the explanations given in a paper should be sufficient. Open-source should not play any role in the review process.
- Will research that does not come bundled with open-source code be relevant in 10 years?
- Should MIR students learn open-source coding skills to increase their job prospects?

We agreed that source code has a different degree of importance in different subareas of MIR research. In some areas of MIR research (e. g., deep learning) the line between research and software engineering is blurring. As a community, we should maintain a broad view of these issues accepting people on either side of the debate. A further debate may be informed by a recent publication on the topic of open-source software for MIR research, see [1].

References

- 1 Brian McFee, Jong Wook Kim, Mark Cartwright, Justin Salamon, Rachel M. Bittner, Juan Pablo Bello. *Open-Source Practices for Music Signal Processing Research: Recommendations for Transparent, Sustainable, and Reproducible Audio Research*. IEEE Signal Processing Magazine, 36(1), 2019, pp. 128–137.

4.4 Singing Assessment and Performance Evaluation

Participants of Dagstuhl Seminar 19052

License © Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

The role of music technology in music education is visible, but not as much as one may guess or hope for. The reason for this gap may be attributed to the fact that MIR and music education are two different communities with their distinct outlooks and goals. For example, while singing teachers are likely to believe that abstract aesthetic judgments play the major role in their assessment of learners, MIR researchers often assume that the rapid improvements in automatic extraction of pitch, intonation, rhythm, and dynamics from audio signals herald the coming of powerful teaching tools [1]. In this working group, we reflected on such issues and discussed the potential and limitations of specific technologies for music education. The following list gives an impression on the topics covered:

- Slowing down of recorded musical pieces for superior listening and understanding has been a prevalent but technologically simple tool.
- Separation of sources in a mixed recording in order to enable easy listening to a chosen part, or to eliminate a part in order to use the accompaniment to practice along with, has also been sought after. While source separation technology is not robust enough for use with arbitrarily generated mixes, special recordings have been used to facilitate this.
- Automatic feedback may be a boon to self-learning and practice where learners do not have to rely on their perception to guide them. Computer feedback can also be preferred at times as being less intrusive. The type of feedback that can be reliably provided is

about accuracy in pitch and rhythm concerning a reference (i. e., learning by imitation scenario). Some karaoke apps do precisely this based on a very basic transcription of the reference song. Given the difficulty in characterizing the pitch of a note given the continuously varying pitch in the singing of lyrics and the presence of consonants, such feedback is based on some gross measures only. Even feedback on onsets is not completely reliable given the variety of sung phonemes and singing styles.

- Instead of explicit feedback, it may be preferred to provide a visual representation of the detected continuous pitch superposed on the expected melody including events such as vibrato as in the MiruSinger singing interface [2]. Even scoring an imitation is not easy but such visual comparisons can help a large category of learners who aim to imitate chosen reference songs as closely as possible.
- To make learning by imitation easier, it may help further to create audio tracks that render the reference song in the learner's pitch range and with similar voice quality.
- At the opposite extreme, it would be interesting to predict the actual ratings of human experts via regression methods. This might be expected to involve a more holistic examination of the audio recording with possibly multiple musical attributes taken into account simultaneously. Considering that there are probably several distinct ways to sound right, it may be better to focus on identifying no-gos and designing algorithms to trap such forbidden events in the singing.
- Apart from pitch and timing, vocal dynamics and voice texture or timbre are of great interest, especially to the more advanced singers and performers. Expressiveness in performance, for example, owes itself to intonation, dynamics and voice quality. Again, visual representations of these parameters that facilitate comparisons with a reference may be the best option.
- Recently, feedback in the context of choir singing has been of much interest. Choir directors can benefit from automatic detection of dissonance and of tuning drift with time. Singers can also benefit from source separation tools and the ability to get automatic accompaniment of the other parts in order to practice choir singing away from the choir.
- Non-Western music presents distinct challenges based on the specific genre. In classical Indian music, the rendering of transitions between notes and of ornaments is as important as note accuracy. Salient features are extracted from complex pitch movements to create a perceptually meaningful space for comparisons [3]. More research is needed to understand what constitutes out-of-tune singing in this context and how other attributes such as loudness/dynamics and vocal timbre might play a role.

References

- 1 Christian Dittmar, Estefanía Cano, Jakob Abeßer, and Sascha Grollmisch. *Music information retrieval meets music education*. In Dagstuhl Follow-Ups (Vol. 3). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2012, pp. 95–120.
- 2 Tomoyasu Nakano, Masataka Goto, and Yuzuru Hiraga. *MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data*. Proceedings of the IEEE International Symposium on Multimedia Workshops (ISMW), Beijing, 2007, pp. 75–76.
- 3 Chitralakha Gupta and Preeti Rao. *Objective Assessment of Ornamentation in Indian Classical Singing*. In *Speech, Sound and Music Processing: Embracing Research in India*. Springer, Berlin & Heidelberg, Germany, 2011, pp. 1–25.

4.5 Subjective Evaluation and Objective Metrics for Singing Voice Generation and Separation

Participants of Dagstuhl Seminar 19052

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

In light of recent deep learning approaches based on generative models in the fields of source separation and voice synthesis, it is important to discuss an evaluation strategy that is standardized and well accepted. In recent years, there has been some discussion [3] on the effectiveness of objective and perceptually inspired metrics such as SDR, SAR and SIR [1, 2]. For voice synthesis, it has been shown that objective measures like the Mel-Cepstral Distortion do not correlate well with subjective evaluation via listening tests [4]. In this working group, we addressed these issues and proposed the idea of formulating guidelines for publishing evaluation results. In particular, we agreed on the following suggestions:

- Existing objective metrics should always be provided.
- Additionally, subjective evaluations should be conducted.
- In case of a discrepancy between objective metrics and subjective evaluations, a preliminary explanation should be provided, along with supportive and exceptional examples, possibly via a supplementary website.
- GitHub and/or Zenodo should be used for sharing sound examples of the results. Furthermore, if possible, source code or detailed explanations on how the examples were generated should be provided.

With consideration to these guidelines for authors, we also considered the possibility of finding effective objective metrics which can bridge the gap between subjective and objective evaluation. While this seems a daunting task, the possibility of using discriminators based on generative adversarial networks (GANs) was discussed.

References

- 1 Emmanuel Vincent, Rémi Gribonva, and Cédric Févotte. *Performance Measurement in Blind Audio Source Separation*. IEEE Transactions on Audio, Speech and Language Processing, 14(4), pp. 1462–1469.
- 2 Emmanuel Vincent. *Improved Perceptual Metrics for the Evaluation of Audio Source Separation*. International Conference on Latent Variable Analysis and Signal Separation. Springer, Berlin & Heidelberg, Germany, 2012, pp. 430–437.
- 3 Dominic Ward, Hagen Wierstorf, Russell D. Mason, Emad M. Grais, and Mark D. Plumbley. *BSS Eval or PEASS? Predicting the Perception of Singing-Voice Separation*. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, 2018, pp. 596–600.
- 4 Merlijn Blaauw and Jordi Bonada. *A Neural Parametric Singing Synthesizer*. Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech), Stockholm, Sweden, 2017, pp. 4001–4005.
- 5 Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. *Generative Adversarial Nets*. Advances in neural information processing systems, Montreal, Canada, 2014, pp. 2672–2680.

4.6 Symbolic Domain Melody Estimation

Participants of Dagstuhl Seminar 19052

License  Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

In this working group, we considered the task of identifying the melody notes from single-track (e. g., piano) or multi-track symbolic music representation. For melody estimation from multi-track music, this also involves identifying the musical track that contains the melody for each time instance. We call this task “symbolic-domain melody estimation.” To our knowledge, this is a task that has been only sporadically addressed in the literature, and the focus of recent work is mostly on voice separation (i. e., the separation of symbolic music into perceptually independent streams of notes such as voices or melodic lines) [1, 2, 4, 5, 6].

There are a few possible reasons why the task has not received full attention in the community. First, over the past two decades, most research activities in the MIR community are concerned with the analysis, processing, and retrieval of audio-based rather than symbolic representations [3]. Second, for people working with symbolic domain data, melody estimation is typically assumed as a pre-processing step rather than the primary research focus; people tend to assume that the melody is given and work on something else that interests them more. There is no strong demand, or use case, for symbolic-domain melody estimation.

We found a new use case for symbolic-domain melody estimation: the support AI-based music generation. Currently, most research on music generation is on melody/chord generation. Here, symbolic data formats such as the ABC format [7] or MusicXML of lead sheets [8], where the melody track is specified, are used. A natural extension of such research is to generate the accompaniment, which may be composed of multiple musical tracks other than the melody and chords. To learn to generate accompaniment tracks, one can use MIDI files as the training data, as done by Dong et al. [9]. However, as people typically do not specify which track in a MIDI file is the melody track, the MuseGAN model presented by Dong et al. [9] can generate only the accompaniment, not the combination of melody and accompaniment. Liu et al [10] attempted to address this issue by using chord-related representations to connect lead sheets and MIDI files (since we can compute chord-related features from both data). The model can generate not only the melody but also the arrangement for the music of 4-bar long. However, it falls short when it comes to generating more extended music, as that requires paired data of lead sheets and MIDI files that are both long enough.

We proposed that symbolic domain melody estimation, in particular identifying the melody track from a multi-track MIDI file, can be an essential building block toward generating multi-instrument music. With symbolic domain melody estimation, we can learn better the relationship between melody and accompaniment, without the need for paired data of lead sheets and MIDI files.

We also proposed that a possible approach to encourage research along this line is to build a MIDI data set with the melody tracks labeled (for example, from the Lakh MIDI data set [11]). This dataset can then be used to train a supervised model for symbolic domain melody estimation. With the labeled MIDI files, we can evaluate how well unsupervised, rule-based methods for symbolic domain melody estimation works, and whether there is gain by combining supervised and unsupervised methods.

Symbolic domain melody estimation is an interesting problem on its own, and it has strong applications in music generation. We hope this extended abstract can call for more attention toward this research problem.

References

- 1 David Rizo, Pedro J. Ponce de León, Antonio Pertusa, Carlos Pérez-Sancho, and José Manuel Iñesta Quereda. *Melody Track Identification in Music Symbolic Files*. Proceedings of the International Florida Artificial Intelligence Research Society Conference, Melbourne Beach, USA, 2006, pp. 254–259.
- 2 Dimitrios Rafailidis, Alexandros Nanopoulos, Yannis Manolopoulos, and Emilios Cambouropoulos. *Detection of Stream Segments in Symbolic Musical Data*. Proceedings of the International Conference on Music Information Retrieval (ISMIR), Philadelphia, USA, 2008, pp. 83–88.
- 3 Meinard Müller. *Fundamentals of Music Processing – Audio, Analysis, Algorithms, Applications*. Springer Verlag, 2015.
- 4 Patrick Gray, Razvan C. Bunescu. *A Neural Greedy Model for Voice Separation in Symbolic Music*. Proceedings of the International Conference on Music Information Retrieval (ISMIR), New York City, USA, 2016, pp. 782–788.
- 5 Andrew McLeod and Mark Steedman. *HMM-Based Voice Separation of MIDI Performance*. Journal of New Music Research, 2016, 45(1), pp. 17–26.
- 6 Wei-Tsung Lu and Li Su. *Deep Learning Models for Melody Perception: An Investigation on Symbolic Music Data*. Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Honolulu, USA, 2018, pp. 1620–1625.
- 7 Bob L. Sturm, João Felipe Santos, Oded Ben-Tal, and Iryna Korshunova. *Music Transcription Modelling and Composition Using Deep Learning*. Proceedings of the Conference on Computer Simulation of Musical Creativity, 2016.
- 8 Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang. *MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation*. Proceedings of the International Society for Music Information Retrieval Conference, Suzhou, China, 2017, pp. 324–331.
- 9 Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. *MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment*. Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018, pp. 34–41.
- 10 Hao-Min Liu and Yi-Hsuan Yang. *Lead Sheet Generation and Arrangement by Conditional Generative Adversarial Network*. Proceedings of the International Conference on Machine Learning and Applications (ICMLA), Orlando, USA, 2018, pp. 722–727.
- 11 Colin Raffel. *Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching*. Ph.D. Thesis, Columbia University, 2016.

4.7 Transcription beyond Western Music

Participants of Dagstuhl Seminar 19052

License © Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

In this working group, we discussed aspects related to transcribing and representing music beyond standard Western music notation. Before defining how such a representation should look, it is essential to consider the intended use case.

One use case could be ethnomusicological research and archival purposes. For example, in the context of Georgian vocal music [1], we discussed a software tool for displaying a piano-roll representation, with further visual cues, like Gaussian approximations of sung pitches, annotations for sung lyrics, and so on. In this scenario, many important issues remain. First, it is unclear how to enrich such representations with timbre or tuning attributes. Annotations

for lyrics are very difficult to obtain for this kind of music. The automatic alignment of audio and lyrics would be of great help. An idea that emerged out of this working group was to approach this problem through clustering techniques of MFCC features [2].

Another important use case for music representations is performance practice. We agreed that representations for such a purpose have to be similar to Western score representations to a certain extent. The reason is that musicians are well-trained in reading this notation. Thus, all attempts of replacing such standardized representations did not succeed in the history of music. In general, performance practice representations need to be much more abstract than representations for archival purposes.

We also discussed several historical and modern representations of music and prosody, like neumes, a web representation for Indian classical music⁹, rough categories of sung notes in Western music [3], Japanese music¹⁰, ancient Chinese music¹¹, contemporary music notation [4], conventions for transcribing and annotating the prosody of speech¹², the Music Encoding Initiative [5], and the linking of performance data with score information [6]. A general treatment of music notation is given by Read [7].

Furthermore, we also discussed how to technically approach the problem of creating such representations. One of the most important questions is how reliable the extraction of fundamental frequency (F0) trajectories from audio recordings is. Also, the segmentation of F0 trajectories into notes is a difficult problem. A further aspect is the automatic detection of breathing sounds in singing voice music [8].

In summary, the participants agreed on a great interest in generalized music representations, which lie in between F0 trajectories and note representations.

References

- 1 Frank Scherbaum, Meinard Müller, and Sebastian Rosenzweig. *Rechnergestützte Musikethnologie am Beispiel historischer Aufnahmen mehrstimmiger georgischer Vokalmusik*. Proceedings of the GI Jahrestagung, 2017.
- 2 Ben Gold, Nelson Morgan, and Dan Ellis. *Speech and Audio Signal Processing*. John Wiley & Sons, Hoboken, USA, 2011.
- 3 Jiajie Dai. *Modelling Intonation and Interaction in Vocal Ensembles*. Ph.D. thesis, Queen Mary University of London, to appear.
- 4 Jonathan Feist. *Berklee Contemporary Music Notation*. Berklee Press, Boston, USA, 2017.
- 5 Andrew Hankinson, Perry Roland, and Ichiro Fujinaga. *The Music Encoding Initiative as a Document-Encoding Framework*. Proceedings of the International Society for Music Information Retrieval Conference, Miami, USA, 2011, pp. 293–298.
- 6 Johanna Devaney and Hubert Léveillé Gauvin. *Representing and Linking Music Performance Data with Score Information*. Proceedings of the International Workshop on Digital Libraries for Musicology, New York City, USA, 2016, pp. 1–8.
- 7 Gardner Read. *Music Notation*. Crescendo Publishing, Boston, USA, 1964.
- 8 Tomoyasu Nakano, Jun Ogata, Masataka Goto, and Yuzuru Hiraga. *Analysis and Automatic Detection of Breath Sounds in Unaccompanied Singing Voice*. Proceedings of the International Conference on Music Perception and Cognition, Sapporo, Japan, 2008, pp. 387–390.

⁹ <https://autrimncpa.wordpress.com>

¹⁰ In this tradition, extra symbols exist that indicate how to express lyrics, called “Shigin.”

¹¹ There exists a traditional musical notation method called “Gongche” notation.

¹² A set of conventions is called tones and break indices (ToBI).

4.8 Demo Session

Participants of Dagstuhl Seminar 19052

License © Creative Commons BY 3.0 Unported license
© Participants of Dagstuhl Seminar 19052

On Thursday evening, we had a demo and late-breaking news session. The contributions included scientific ideas, graphical user interfaces, data sets, and audio examples. The following list gives an overview of the contributions:

- Sebastian Schlecht, Sebastian Rosenzweig, Christof Weiß: Realtime Dissonance Detection
- Ryo Nishikimi: Automatic Singing Transcription
- Zhiyao Duan: AIR Lab Demo
- Yi-Hsuan Yang: Song Mixer
- Yi-Hsuan Yang: Latent Inspector
- Justin Salamon: Crepe
- Estefanía Cano Cerón: Colombian Music Archive
- Pritish Chandna: Generative Models for Singing Voice Synthesis
- Simon Dixon: Tony
- Juhan Nam: Piano Re-Performance
- Juhan Nam: VirtuosoNet: Expressive Piano Performance Rendering
- Juhan Nam: Singing Voice Synthesis Using Conditional GAN

5 Music and Recording Sessions

As one major overall topic, polyphonic singing was discussed over the course of the seminar. In particular, participants mentioned their interest in Western choral singing in several stimulus talks and submitted abstracts. In this context, pitch, intonation, and singer interaction constitute essential aspects. Detailed studies on such aspects typically require multitrack recordings comprising one or several tracks per singer as well as manual annotations, e. g., in terms of an aligned musical score and lyrics. However, recording multitrack audio in the choir context is challenging, since singers can hardly be recorded in separation and conventional microphones suffer from bleeding between different voices. The lack of suitable recordings and the joy of singing led to the idea of forming a small choir of Dagstuhl participants in order to record a multitrack dataset of choir singing using different types of microphones and sensors that capture both individual voices and the overall acoustic impression.

5.1 Choir Rehearsals

Christof Weiß (Universität Erlangen-Nürnberg, DE), Sebastian Rosenzweig (Universität Erlangen-Nürnberg, DE), Helena Cuesta (UPF – Barcelona, ES), Frank Scherbaum (Universität Potsdam – Golm, DE), Emilia Gómez (UPF – Barcelona, ES), Meinard Müller (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Christof Weiß, Sebastian Rosenzweig, Helena Cuesta, Frank Scherbaum, Emilia Gómez, and Meinard Müller

In several rehearsals (see Figure 1) taking place during breaks and in the evenings, we tried a number of choir pieces collected in advance of the seminar. After several tries, we selected Anton Bruckner’s Gradual “Locus iste” (WAB 23) in Latin (Figure 2). This 3-minute long choir piece is musically interesting, contains several melodic and harmonic challenges, and covers a large part of each voice’s tessitura. Beyond that, this piece is part of the *Choral singing dataset* [1], thus allowing for interesting comparative studies. As further works, we considered the piece “Tebe poem” by the Bulgarian composer Dobri Hristov, the Catalan traditional song “El rossinyol,” and “Otche nash” by the Russian composer Nikolai Kedrov, among others. All pieces were written for SATB choir in four parts, each of which we could perform with at least two singers (see Table 1 for details). The rehearsals usually started with a vocal warm-up led by Polina Proutskova, who works as a singing teacher. Composer Christof Weiß took the role of the choir director for practicing and performing the pieces. Over the first three days of the seminar, the choir steadily improved and reached a reasonable level of musical quality, which might be considered representative of a good amateur choir. Despite the varying level of singing training, the choir could produce a quite homogeneous sound.

The recording session mostly focused on “Locus iste” and included different variations of singer configuration and expressions. Beyond several runs performed by the full choir, we also recorded this piece in two different quartet versions with only one singer for each part (Table 1). For studying the effect of unison singing, we captured the beginning part with the basses only, thereby varying the number of singers (1–5 bass singers). Additionally, we recorded “Tebe poem” with the full choir and a small number of intonation exercises [2] with one of the quartets (*Quartet II*).



■ **Figure 1** Choir rehearsal.

Allegro moderato

p *mf* *f* *p*

Lo - cus i - ste a De - o fa - ctus est, lo - cus i - ste a De - o fa - ctus est, a De - o, De - o fa - ctus est

Lo - cus i - ste a De - o fa - ctus est, lo - cus i - ste a De - o fa - ctus est, a De - o, De - o fa - ctus est

Lo - cus i - ste a De - o fa - ctus est, lo - cus i - ste a De - o fa - ctus est, a De - o, De - o fa - ctus est

Lo - cus i - ste a De - o fa - ctus est, lo - cus i - ste a De - o fa - ctus est, a De - o, De - o fa - ctus est

■ **Figure 2** First measures of Anton Bruckner’s “Locus iste” (WAB 23).

References

- 1 Helena Cuesta, Emilia Gómez, Agusín Martorell, and Felipe Loáiciga. *Analysis of Intonation in Unison Choir Singing*. Proceedings of the International Conference on Music Perception and Cognition (ICMPC), 2018.
- 2 Per-Gunnar Alldahl. *Choral Intonation*. Gehrman, Stockholm, Sweden, 2008.

5.2 Recording Documentation

Sebastian Rosenzweig (Universität Erlangen-Nürnberg, DE), Helena Cuesta (UPF – Barcelona, ES), Frank Scherbaum (Universität Potsdam – Golm, DE), Christof Weiß (Universität Erlangen-Nürnberg, DE), Emilia Gómez (UPF – Barcelona, ES), Meinard Müller (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Sebastian Rosenzweig, Helena Cuesta, Frank Scherbaum, Christof Weiß, Emilia Gómez, and Meinard Müller

The recording session took place on the third day of the seminar in room “Kaiserslautern.” The room was equipped with recording devices, several microphones and many cables, organized and operated by Sebastian Rosenzweig, Helena Cuesta, and Frank Scherbaum. To precisely study the individual singers’ performances and behaviors, we recorded selected singers with

■ **Table 1** Singers participating in the choir recordings.

| | <i>Quartet I</i> | <i>Quartet II</i> | <i>Full choir (additional singers)</i> |
|----------------|------------------|-------------------|--|
| <i>Soprano</i> | Emilia Gómez | Polina Proutskova | |
| <i>Alto</i> | Rachel Bittner | Cynthia Liem | |
| <i>Tenor</i> | Justin Salamon | Simon Dixon | Zhiyao Duan, Tomoyasu Nakano |
| <i>Bass</i> | Meinard Müller | Frank Zalkow | Frank Kurth, Sebastian Schlecht, Li Su |



■ **Figure 3** Microphone setup for three singers.

multiple close-up microphones such as throat, headset, and dynamic microphones. Throat microphones, which capture the vibrations of a singer’s throat, have shown to be particularly useful for such studies thanks to their robustness to cross-talk from other singers [1]. The microphone setup for the singers is demonstrated in Figure 3. With this setup, we recorded two singers per voice section in the full choir and each singer in the quartets. Additionally, we recorded each setting with a stereo room microphone placed in a distance of about three meters from the singers to capture the overall impression.

In parallel to the multichannel audio recordings, we documented the session with videos and pictures. Furthermore, we equipped one singer with an ambulatory monitoring system to acquire behavioral and physiological data during the performances.¹³ Another singer was equipped with binaural microphones to record the choir performances from a singer’s perspective.

After the seminar, we plan to collect, synchronize, cut, and annotate the recorded material with the goal to create a publicly accessible dataset that may serve various research purposes for melody and voice processing. All singers already provided their consent to release the dataset along with all metadata and annotations for research.

References

- 1 Frank Scherbaum. *On the Benefit of Larynx-Microphone Field Recordings for the Documentation and Analysis of Polyphonic Vocal Music*. Proceedings of the International Workshop on Folk Music Analysis, Dublin, Ireland, 2016, pp. 80–87.

¹³<http://www.vu-ams.nl/>

Participants

- Rachel Bittner
Spotify – New York, US
- Estefanía Cano Cerón
Fraunhofer IDMT, DE & A*STAR – Singapore, SG
- Michèle Castellengo
Sorbonne University – Paris, FR
- Pritish Chandna
UPF – Barcelona, ES
- Helena Cuesta
UPF – Barcelona, ES
- Johanna Devaney
Brooklyn College, US
- Simon Dixon
Queen Mary University of London, GB
- Zhiyao Duan
University of Rochester, US
- Emilia Gómez
UPF – Barcelona, ES
- Masataka Goto
AIST – Tsukuba, JP
- Frank Kurth
Fraunhofer FKIE – Wachtberg, DE
- Cynthia Liem
TU Delft, NL
- Antoine Liutkus
INRIA, University of Montpellier, FR
- Meinard Müller
Universität Erlangen-Nürnberg, DE
- Tomoyasu Nakano
AIST – Tsukuba, JP
- Juhan Nam
KAIST – Daejeon, KR
- Ryo Nishikimi
Kyoto University, JP
- Geoffroy Peeters
Telecom ParisTech, FR
- Polina Proutskova
Queen Mary University of London, GB
- Preeti Rao
Indian Institute of Technology Bombay, IN
- Sebastian Rosenzweig
Universität Erlangen-Nürnberg, DE
- Justin Salamon
Adobe Research, US
- Frank Scherbaum
Universität Potsdam – Golm, DE
- Sebastian J. Schlecht
Universität Erlangen-Nürnberg, DE
- Li Su
Academia Sinica – Taipei, TW
- Tomoki Toda
Nagoya University, JP
- Julián Urbano
TU Delft, NL
- Anja Volk
Utrecht University, NL
- Ye Wang
National University of Singapore, SG
- Christof Weiß
Universität Erlangen-Nürnberg, DE
- Yi-Hsuan Yang
Academia Sinica – Taipei, TW
- Frank Zalkow
Universität Erlangen-Nürnberg, DE

