



DAGSTUHL REPORTS

Volume 6, Issue 4, April 2016

Analysis, Interpretation and Benefit of User-Generated Data: Computer Science Meets Communication Studies (Dagstuhl Seminar 16141) <i>Thorsten Quandt, German Shegalov, Helle Sjøvaag, and Gottfried Vossen</i>	1
Multidisciplinary Approaches to Multivalued Data: Modeling, Visualization, Analysis (Dagstuhl Seminar 16142) <i>Ingrid Hotz, Evren Özarslan, and Thomas Schultz</i>	16
Foundations of Data Management (Dagstuhl Perspectives Workshop 16151) <i>Marcelo Arenas, Richard Hull, Wim Martens, Tova Milo, and Thomas Schwentick</i>	39
Tensor Computing for Internet of Things (Dagstuhl Perspectives Workshop 16152) <i>Evrım Acar, Animashree Anandkumar, Lenore Mullin, Sebnem Rusitschka, and Volker Tresp</i>	57
Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments (Dagstuhl Seminar 16161) <i>Elena Cabrio, Graeme Hirst, Serena Villata, and Adam Wyner</i>	80
Managing Technical Debt in Software Engineering (Dagstuhl Seminar 16162) <i>Paris Avgeriou, Philippe Kruchten, Ipek Ozkaya, and Carolyn Seaman</i>	110
Algorithmic Methods for Optimization in Public Transport (Dagstuhl Seminar 16171) <i>Leo G. Kroon, Anita Schöbel, and Dorothea Wagner</i>	139
Machine Learning for Dynamic Software Analysis: Potentials and Limits (Dagstuhl Seminar 16172) <i>Amel Bennaceur, Dimitra Giannakopoulou, Reiner Hähnle, and Karl Meinke</i>	161

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

October, 2016

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
- Stephan Diehl
- Hans Hagen
- Hannes Hartenstein
- Oliver Kohlbacher
- Stephan Merz
- Bernhard Mitschang
- Bernhard Nebel
- Bernt Schiele
- Nicole Schweikardt
- Raimund Seidel (*Editor-in-Chief*)
- Arjen P. de Vries
- Klaus Wehrle
- Reinhard Wilhelm

Editorial Office

Marc Herbstritt (*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.6.4.i

Report from Dagstuhl Seminar 16141

Analysis, Interpretation and Benefit of User-Generated Data: Computer Science Meets Communication Studies

Edited by

Thorsten Quandt¹, German Shegalov², Helle Sjøvaag³, and
Gottfried Vossen⁴

1 Universität Münster, DE, thorsten.quandt@uni-muenster.de

2 Twitter – San Francisco, US, gera@twitter.com

3 University of Bergen, NO, helle.sjovaag@uib.no

4 Universität Münster, DE, vossen@wi.uni-muenster.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16141 “Analysis, Interpretation and Benefit of User-Generated Data: Computer Science Meets Communication Studies”.

Seminar April 3–8, 2016 – <http://www.dagstuhl.de/16141>

1998 ACM Subject Classification H. Information Systems

Keywords and phrases Communication Studies, Text Analysis, NLP, Text Mining, Topic Detection, Sentiment Analysis, Machine Learning

Digital Object Identifier 10.4230/DagRep.6.4.1

1 Executive Summary

Thorsten Quandt

German Shegalov

Helle Sjøvaag

Gottfried Vossen

License © Creative Commons BY 3.0 Unported license
© Thorsten Quandt, German Shegalov, Helle Sjøvaag, and Gottfried Vossen

The success of the Internet as a communication technology and tool for human interaction in countless contexts, including production and trade, has had a dramatic impact on modern societies. With diffusion rates nearing one hundred percent in most societal groups, there is virtually no one whose life is not influenced by online communication – either directly or indirectly. Every day, private end users and business users act and interact online, producing immense amounts of data. Many disciplines, including computer science, computer linguistics, psychology, and communication studies, have identified ‘big data’ generated by online users as a research field. As a result, big data has become a somewhat over-hyped catch-all term for many different types of data, which are analyzed through varying methods for multiple purposes. This ranges from an analysis of (unstructured) Twitter or Facebook content to rule-structured texts as found in the professional media (i.e., news websites). The implication of value generated through sheer size of data sets is misleading, though – much of the value is based on the nature of these data sets as being user-generated, either on purpose or inevitably (and often unknowingly) as behavioral traces of actions with divergent aims.



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

Analysis, Interpretation and Benefit of User-Generated Data: Computer Science Meets Communication Studies, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 1–15

Editors: Thorsten Quandt, German Shegalov, Helle Sjøvaag, and Gottfried Vossen



DAGSTUHL
REPORTS Dagstuhl Reports

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

Big data sets generated by human users pose some challenges to the scientific disciplines that are interested in them: Typically, computer scientists have the knowledge and tools to access, extract and process big data sets. However, the analysis and interpretation of such data mirrors the interactions of users who produced the data and is not following a purely technological logic. In other words, such data has a human/social component, and in order to interpret and understand it, social-scientific theories and methods are helpful. Social scientists, however, typically do not specialize in the practicalities of online technologies and of programming. While they have theoretical approaches and empirical methods available that can be helpful in the analysis of user generated content – and this is especially true for communication scholars who specialize in the analysis of (online) media content –, their possibilities to access and process data are limited (as this is not core to their field yet).

Consequently, both disciplinary approaches will not be able to fully address the challenges of analyzing big data based on user (inter)action from the perspective of their own ‘silo’. A combination of the two approaches seems fruitful, as each discipline may help in solving the problems of the other, and the sum will be more than its parts – leading to a better understanding of social interaction and human communication in a digitized world. This seminar will bring together both computer scientists interested in the analysis of (large-scale) user-generated data, and communication scholars interested in computer-assisted acquisition and processing of such data. It is intended to start a fruitful dialogue on potential approaches, methods, uses and benefits of a cooperation between the two disciplines, and it will also include the input of practitioners in the field of media and business who will offer valuable insights into practical use cases.

2 Table of Contents

Executive Summary

Thorsten Quandt, German Shegalov, Helle Sjøvaag, and Gottfried Vossen 1

Working groups

Analyzing Text Microstructures
Christian Baden and Tatjana Scheffler 4

Visions for the Computer-Assisted Identification, Analysis and Evaluation of Texts
Christian Baden 6

Interactions between Computer Science (CS) and Journalism (Studies) in the Future
David Domingo, Johann-Christoph Freytag, Ari Heimonen, and Rodrigo Zamith 7

What is not there yet? Dreams and Visions
Martin Emmer, Elisabeth Günther, Wiebke Loosen, Alexander Löser, and Gottfried Vossen 9

Methods on Obtaining Curated Data (such as for trend detection, or for understanding social problems, for power-law distributed data)
Alexander Löser 10

Funding Workshop
Helle Sjøvaag 10

Workshop on Data Journalism
Helle Sjøvaag 11

Workshop on Methods
Helle Sjøvaag and Thorsten Quandt 12

Workshop on Relation Analysis
Helle Sjøvaag 12

Workshop on What is not there yet?
Helle Sjøvaag 13

Panel discussions

Data Access
Thorsten Quandt 14

Participants 15

3 Working groups

3.1 Analyzing Text Microstructures

Christian Baden (The Hebrew University of Jerusalem, IL) and Tatjana Scheffler (Universität Potsdam, DE)

License  Creative Commons BY 3.0 Unported license
© Christian Baden and Tatjana Scheffler

In this working group session, we discussed three major challenges in analyzing text microstructures.

Challenge 1: Finding frames

- We use the same techniques for allegedly theoretically different purposes: LDA “finds frames” or “finds topics” depending on what we want, but really it is one technique, so this does not make sense. Frames and topics should generally be orthogonal, such that one topic can take different frames, and one frame can take different topics.
- There are existing strategies to run a topic extraction tool, then control for that, and call all further patterns frames; this works but is theoretically unsatisfying. Also, other types of structures have been analyzed in this way (speech acts, event structure, etc.). Frames have a specific theoretical structure (four frame elements, following Entman):

Evaluation
|
Cause - Focal Concern - Projection/Treatment

- We discussed an idea based on this structure – use this structure to construct frames from texts:
 1. use topical text contents (headlines, lead paragraphs) to identify a (narrow) topic
 2. structure remaining textual contents
 3. use syntax, sequence, connectors, grammatical information (whatever is useful/available) to figure out which other contents are related how to the focal topic

Challenge 1b: Focusing topic models (or other similar techniques)

- Models tend to perform better if more textual content is excluded (even up to excluding verbs).
- Possible cause seem to be stylistic differences.
- This is also highly dependent on the size of documents one analyses – aggregating individual texts into larger “documents” leads to different results in topic models.
- Such insights might help focus pattern finding algorithms on different text properties: topical (if highly reduced), stylistic, etc.
- There are also formal solutions, where one can run topic models in a multilevel logic: annotating authors, they can be trained to disregard author-specific variation and focus on content differences; and there is no reason why one can do this only for partialing out author-related variance, might be a great strategy for comparative research.

Challenge 2: Similarity of texts

- Determining the similarity of text has many uses: deduplication, detecting taken-over materials, assessing diversity, etc.; however, many existing approaches are unclear about what exactly they mean, as they depend on features of the text whose theoretical relevance is heterogeneous and/or unclear.

- Four basic types of similarity that are of interest: literally identical sequences (quotes, plagiarism, unedited text); content similarity (the same topic and arguments), stylistic (the same way of expressing content, sentence style, etc), and sentiment.
- Furthermore, it may be relevant to assess similarity despite different languages; there is rich work in machine translation, evaluating translation quality based on these above dimensions, and also providing avenues for interlingual comparison; however, machine translation increases comparison error in ways not yet sufficiently understood.
- Evaluation methods developed for machine translation (e.g., similarity to a (set of) reference translation(s)) may also be used in a monolingual setting to determine the similarity of sentences/texts.
- Problems related to challenge 2:
 - (1) Matching (finding out which of very many possible pairs are worth comparing/have relevant similarity); for this, using meta data may be useful to consider only likely combinations (date, for instance). Parallelization is probably desirable, to avoid memory problems. In addition, internal text structure (e.g., sections, zoning, substructure of journalistic texts) can be used to presegment documents for comparison of smaller chunks. Possible detection problems:
 - Some kinds of texts (e.g., sports reporting) are structurally very similar because there are just few ways of saying something; can also be considered valid result though.
 - Similarity often does not concern entire texts, but a text may be similar to a part of another text only, or both only overlap in particular paragraphs; so text-level similarity scores might be too crude
 - However, there is a specific interest in determining what are possible elaborations/truncations of texts, so both determining if there are similar passages, and determining what is different are important.
 - The shorter the text the more likely does one find similar others just by chance.
 - (2) Measuring/scoring the similarity. For each kind, there are good algorithms existing that can be developed and applied: for literal identity, plagiarism checkers; for contentual similarity, comparisons of extracted entities (rank sum, bags of words strategies, Jaccard, etc); for stylistic, bag of words & linguistic resources; for sentiment, new generation sentiment measures that take into account differentiated scores and intensifiers/negations, also some machine learning approaches. There may be a point in keeping different kinds of similarity apart and finding typical patterns of similarity (e.g., high contentual low literal/stylistic similarity → paraphrase, some literal, low sentiment → quote/challenge, etc.).

Challenge 3: Metaphors

- Metaphors matter both for sentiment (they have evaluative implications) and for framing (they structure content), but are difficult to find. Existing approaches are deductive, domain specific, and laborious, and they still detect a lot of cases that are actually literal uses, not metaphors.
- One idea to solve this: Use related content (e.g., Wikipedia, dictionaries) to determine if a word that might be a metaphor is used in a context related to its definition, or out of such context (so it is probably metaphorical).

3.2 Visions for the Computer-Assisted Identification, Analysis and Evaluation of Texts

Christian Baden (The Hebrew University of Jerusalem, IL)

License  Creative Commons BY 3.0 Unported license
© Christian Baden

The group identified three main areas of basic challenges for the computerized treatment of texts:

1. accessing texts, from various archives, via scraping, apis, etc.
2. curating texts/evaluating the quality of repositories (formatting, standardization, annotation/metadata, api transparency, etc.)
3. analyzing text (henceforth the focus of discussion)

The third challenge was discussed as being the one of utmost important, and the group approached it from two angles:

Angle 1: Finding entities & patterns in text

- There is a great potential for automation, however, fully supervised approaches are very labour intensive, while unsupervised approaches are hard to trust/use/get published. In essence, there is an urgent need for transparency (what does it do) and intellegibility (ability to theoretically evaluate the rules).
- One idea that was favored by the group was an alternative process to simplistic automatic content analysis: We propose to use machine learning not to solve the task, but to propose additional indicators and possible additional rules. The process is as follows: (1) start from set of indicators/rules and train ML to find other contents of similar or related kind; (2) generate rules that, if applied, improve performance; (3) return to human in a format that can be understood/evaluated, and if confirmed, integrate into model (iterative semi-supervised approach). This process can be applied to various problems: language fuzziness (find possible typo/variants/synonyms), entity extraction (find additional names), pattern recognition (find additional related components), etc. Another advantage is that the procedure can explicate detection rules, so we can learn not only what is in the text, but also what are underlying structures of discourse useful for analysis.
- The process of finding additional contents of similar kind can also be used to augment/-contextualize/evaluate information. Finding other texts about the same event, or other pictures of the same thing, etc. might be useful to augment journalists' information base, criticize one-sided information, detect contested information, check veracity, etc.
- One further extension is potentially interesting: Like in knowledge graphs, additional information available in discourse can be integrated by linking entities to online resources (e.g., Wikipedia, dictionaries, prior discourse) for elaboration and classification ('active intelligence'/intelligent classification/generation of background information)-

Angle 2: Doing this collaboratively

- There are lots of tools and approaches out there, but little collaboration. This leads to problems in findability of tools, documentation, standardization (of approaches, exchange formats, etc), and also referencing/crediting and related incentive systems. In short, many researchers are solving the same problems in parallel, and make the same mistakes in parallel, instead of working together.
- One idea to solve this is to build an infrastructure that facilitates collaboration (github ++): This should not be only about archiving/sorting/finding tools (possibly with some

mechanism for identifying if existing projects look related to the one you are currently working on, suggesting code/tools), but also about rendering the collaboration and use visible (so one can show that one's tool is useful, gain credit, and get references for developed tools, which makes this worthwhile career-wise).

- Furthermore, and related to this, there is a need for more interdisciplinary education between communication and computer science (iSchools, data science, communication programs that train computational skills/computer science programs that relate to social science research methodology/applications).

Overall, the group came to one central conclusion:

What is needed is not an integrated catchall solution using fancy maths and big red buttons, but an assortment of tools specialized to capture specific ingredients of social scientific concepts, which are well described, allow human intervention, and generate output formats that can be integrated analytically (so, no automated-frame-finder but tools extracting entity classes, relation classes, stylistic contents, etc.).

3.3 Interactions between Computer Science (CS) and Journalism (Studies) in the Future

David Domingo (Free University of Brussels, BE), Johann-Christoph Freytag (HU Berlin, DE), Ari Heinonen (University of Tampere, FI), and Rodrigo Zamith (University of Massachusetts – Amherst, US)

License © Creative Commons BY 3.0 Unported license

© David Domingo, Johann-Christoph Freytag, Ari Heinonen, and Rodrigo Zamith

In this working group session, we discussed various questions on the potential interactions between computer science and journalism (studies). The discussion revealed that some parts of journalistic work can be substituted by computers and robots, some others cannot. However, we found that the discussion about “substitution” is misleading, as the new configurations of information distribution will require both humans and computers, and that it's not about competition, but about new forms of journalistic work.

1. Can computer systems substitute human journalists?
 - (a) If journalism is a filter between events/news and citizens/consumers then an algorithm could do the filtering task. But journalists may still be better filters, based on their intuition, judgments, and a more global view on events and their relationships. Challenge for CS: develop more complex and comprehensive algorithms/methods for performing filtering (almost) like journalists. If computer systems achieve that level of reliable filtering, journalists could focus on other, “more interesting” tasks. Already happening in some areas of journalism, such as weather/sports reports (robot journalism). However, journalists may be more than just filters; they are also “sense makers” that bring information together. In theory, computers may also be able to perform that task.
 - (b) Currently not all events/information on events is digital and available online. Therefore, sensors (such as cameras), which are currently used for surveillance, could also be used for capturing events that can then be evaluated and filtered by humans and/or

machines. Challenge: how to get more sensors integrated into event-generating networks? Risk: event-generating networks might also be used for other purposes than news generation, with a less democratic goal.

- (c) Computers could produce more efficient multi-dimensional news reports that show the information more like a process rather than an “end product”. In this way we can better represent/conservate the complexity of reality and make the process of news report generation more transparent (tracing the provenance of information, see below).
2. Some deeper reflection on Aspect 1: We may not be able to substitute human journalist with computer systems completely with current technology, neither may it be desirable due to possible manipulations of automated systems. Journalist may in any case still be needed as safeguards of the process of news generation.
 3. We can also improve journalistic tasks with computer-based systems, without substituting humans with robots. We developed two ideas in this direction:
 - (a) Enriching news reports with information about the newsgathering process. This could be done by semi-automatically logging actions, documents and sources that are used during that process, thus making it more transparent for consumers and other journalists. One of the practical ways to give access to the newsgathering log data is to link it to individual elements of news reports. (For example, automatically storing a list of the documents a journalist accessed or keywords used in searches, and allowing the journalist to select the trace data to make public.)
 - (b) Improving journalistic memory by better structuring news archives with time series and algorithmic calculations, thus allowing to answer searches and queries with time dimension, showing the evolution of actors, topics, contexts. Example: how has the relationship between Syria and the US changed over the last 20 years? Algorithm could highlight the sentiment in reports of relations between main actors and the topics usually discussed in those reports, presenting it longitudinally as a timeline.
 4. In a more near term it is advantageous/desirable to simplify the interaction between the journalist (or journalism scholar) and the set of computer tools that he/she uses. (Put differently: make the current state of the art more accessible to end-users.) Using the paradigm of SQL as a declarative language, it could be possible for the user to simply express what is the desired outcome of an analytical process, together with possible sources and filters/constraints, thus freeing the users of technical cumbersome details about the algorithms and methods used during that process. (For example: `SELECT NamedEntities FROM doclist AND SENTIMENTANALYZE (NamedEntities. Obama AND NamedEntities.Putin) AND ExtractTopics`. That query would automatically apply an NER tool and Topic Modeling tool to extract information from a set of unstructured documents and save it in as elements in a database.) At the same time, users may have the option of exploring the tasks/steps that the computer system may apply, in order to give more experienced/advanced users the ability to fine-tune the analytical process, or to let new users understand the operations that are being performed.

3.4 What is not there yet? Dreams and Visions

Martin Emmer (FU Berlin, DE), Elisabeth Günther (Universität Münster, DE), Wiebke Loosen (Hans-Bredow-Institut – Hamburg, DE), Alexander Löser (Beuth Hochschule für Technik – Berlin, DE), and Gottfried Vossen (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license

© Martin Emmer, Elisabeth Günther, Wiebke Loosen, Alexander Löser, and Gottfried Vossen

The group discussed several “levels” of visions.

Vision level 1: It would be desirable to make the differences between “forums” of public debate visible: like comment sections of tabloids vs. quality papers, or the papers itself. Furthermore, one might want to look for argument structures, types of authors, audiences etc. The goal here would be a sensor for “public opinion”, delivering data that can be compared to results of public opinion surveys. This would offer further insight in public communication processes.

Vision level 2: The second ‘level’ is lifting the first approach to a more global/macro level, as the group identified one major challenge today: organizing political debates under conditions of extreme speed, heterogeneity, and ambiguity (i.e. fragmentation, filter bubbles, increasing masses of information). So it is not a small set of tv-news and quality papers that organizes this discussion anymore.

Based on this, the overall goals would be a system that analyzes the mediated public sphere online, in order to provide the society – citizens AND elites – with information about the issues, arguments and opinions that are currently debated in society.

Such a system should have various features and functions:

- It should make the public debate “visible” und understandable.
- It may be designed as central platform (liquid democracy) – or maybe as distributed, self-organizing systems?
- The role of state remained unclear. The group agreed that there should be a separation of government and media. So the group argued for a self-organizing system.
- Data protection: It remained disputed what info to include in analysis and presentation of data.
- Public broadcasting: Maybe there could be a new role for this type of actor.
- Fact checking should be included.
- All types of media should be included as well: text, video, pictures.

What form could a project like that have? The group collected various features:

- assistant system for journalists, giving overview over state of debate, facts
- app, usable by everybody
- bot that participates in debates, enriching it
- target groups: reaching highly-involved and less interested citizens at the same time
- low threshold-strategy to get many people using the system
- dealing with problems: user selectivity, instrumental use of results etc.

Vision level 3: Finally, we added a global and ethical dimension, asking: Can such a system be used in multiple countries (Europe)? The discussion revolved around two aspects:

- Would it be useful to build such a system in hardware? This would allow democratic values to be encapsulated in a system. Having independent systems for Europe in order to secure data security would be crucial, then.

- Dystopian pictures of future (science-fiction) remained: Often, we refer to capitalism as the main agent of acceleration and content multiplication. So we asked: Are there possible features to de-accelerate debates?

3.5 Methods on Obtaining Curated Data (such as for trend detection, or for understanding social problems, for power-law distributed data)

Alexander Löser (Beuth Hochschule für Technik – Berlin, DE)

License  Creative Commons BY 3.0 Unported license
© Alexander Löser

The group identified two differing goals for the two disciplinary fields involved in the workshop:

1. Computer studies: Train a smart machine (super intelligence) that does a task (spotting terrorists, products, winning strategies for playing “GO”).
2. Social sciences: Learn from human behavior and abstract it into a report.

Several methods in the two fields were identified by the group:

- A. Observing + Transforming (done in most cases by CS people) This includes text mining from samples (<https://aclweb.org/anthology/>), but also transforming image representation into text, transforming tables into text (robot journalism).
- B. Asking people/Survey methods This includes micro-task crowd sourcing and active learning (for sampling strategies see <http://burrsettles.com/pub/settles.activelearning.pdf>)
- C. Controlled Experiments There is a huge body of research in SS, often ignored by CS, because they set up experiments which are focused around machines. Furthermore, there is a lot of potential in “Games with a purpose”. There are some issues to be solved with that approach, though. One major question is often not answered: “What is the stimulus?” The preferential method here is to eliminate all other factors that might obfuscate the outcome of your experiments. Additional problems may arise from fatigue (one solution may be taking more people, but avoiding long game time).
- D. Simulation This may include creating a machine that is “creatively” creating curated, labeled data (Dynamic programming, Monte Carlo simulations).
- E. Ensemble methods At the end, the major solution may be learning an ensemble from these methods (Boosting, Bagging, DNN:CNN, RNNs or LSTMs?....) and iterate (go back to data sampling and curation) until “good enough”.

3.6 Funding Workshop

Helle Sjøvaag (University of Bergen, NO)

License  Creative Commons BY 3.0 Unported license
© Helle Sjøvaag

The last workshop focused on funding schemes for possible joint future applications. Horizon 2020 and ERC were discussed as EU funding schemes. Other schemes mentioned include EURA (collaboration between certain EU countries); COST Actions (networking scheme); RISE (new funding for research exchange); UNESCO (but this is policy oriented). In the US, foundations are the most likely source of funding, including the Knight Foundation, The

Democracy Fund, Google, The Spence Foundation, TOW Centre, and Reinhold's Journalism Institute. Other funding schemes include the Dutch Press Fund, Tekkis, a Finnish agency. The question of industry funding was also raised.

3.7 Workshop on Data Journalism

Helle Sjøvaag (University of Bergen, NO)

License  Creative Commons BY 3.0 Unported license
© Helle Sjøvaag

The breakout workshop on data journalism met in three sessions during the week. Martin Emmer (FU Berlin, DE), Gottfried Vossen (Universität Münster, DE), Seth C. Lewis (University of Oregon, US), Rodrigo Zamith (University of Massachusetts – Amherst, US), amian Trilling (University of Amsterdam, NL), Ralf Schenkel (Universität Trier, DE), Ari Heinonen (University of Tampere, FI), Jukka Huhtamäki (Tampere University of Technology, FI), Raul Ferrer Conill (Karlstad University, SE) and Helle Sjøvaag (University of Bergen, NO) were the core participants.

Data journalism started as data assisted reporting in the 1960s and has been described as precision journalism or data assisted journalism. It involves journalism practice using social science methods, databases, and using data to do journalism. The data used to do data journalism is typically government data, leaks, and open data. Sometimes this data comes in unmanageable form, such as PDFs or even printouts. Data journalism today is primarily practiced in big newsrooms with the resources to allocate staff to data journalism processes, such as design, data science, statistics and journalism. Because of the nature of the data, data journalism typically involves scraping and visual analytics, and the work frequently requires teamwork.

Challenges to data journalism include acquiring the skills needed to handle tools for data journalism research and presentation. Furthermore, most data journalism projects are ad-hoc projects, with few reproducible workflows. Hence, contingency in data storage and scalable workflow models is a problem. For journalists, the challenge is how to better turn unstructured documents into structured documents. For research the challenge is to look beyond the text as object of study. Data journalism is more than text, which challenges the way we look at societal communication. For journalists and researchers alike, a common challenge is how to treat journalism as data over a large repository beyond archiving, as semantic networks. Part of a solution to this problem is to create transparent workflows.

The discussion in the workshop developed into an effort to establish a collaborative research design for a project on data journalism, based on the interdisciplinarity necessitated by the research object: as visuals, background data, data bases, hidden or licensed data, and text. To study data journalism, the tools as well as the analysis object requires a mixture of social science and computer science approaches. Conceptually, thinking about data journalism in computer science concepts will better facilitate a research design. By using the computer science workflow approach, we then broke the data journalism process into a reference process model, from which the methodology can be derived. And as data journalism is largely about application development, the empirical focus includes both practice (workflow) and 'text' (data).

The data journalism sessions resulted in a rudimentary research design, a collaborative document from which the project can be further developed, and allocation of project leadership.

3.8 Workshop on Methods

Helle Sjøvaag (University of Bergen, NO) and Thorsten Quandt (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license
© Helle Sjøvaag and Thorsten Quandt

This joint session revolved around mapping computer science methods appropriate for communication science research. The methods were divided into three strands:

1. Methods for data access/access to sources/data clearing
2. Language based methods, sentiment analysis, NLP
3. Methods for relation analysis: pattern detection, temporal flows

A collaborative table overviewing the available tools was created. What emerged from the discussion is not only a list of available approaches, but also that most of us using hybrid analysis approaches use the same tools, or the same types of tools. As most researchers need to write their own scripts for scraping, sharing these workflows for static content would be a good idea. The issue of hiring companies to extract the information needed in communication science research was raised, to which the black box problem was discussed.

3.9 Workshop on Relation Analysis

Helle Sjøvaag (University of Bergen, NO)

License  Creative Commons BY 3.0 Unported license
© Helle Sjøvaag

The session included Thorsten Quandt (Universität Münster, DE), David Domingo (Free University of Brussels, BE), Martin Emmer (FU Berlin, DE), Johann-Christoph Freytag (HU Berlin, DE), Jukka Huhtamäki (Tampere University of Technology, FI), Alexander Löser (Beuth Hochschule für Technik – Berlin, DE), Ralf Schenkel (Universität Trier, DE), Gera Shegalov (Twitter – San Francisco, US), Helle Sjøvaag (University of Bergen, NO), Hendrik Stange (Fraunhofer IAIS – St. Augustin, DE), Martin Theobald (Universität Ulm, DE) and Gottfried Vossen (Universität Münster, DE).

The session revolved around analysis of relations between actors in a textual context. For most research in social science, an actor is an individual person or persons that act in a coordinated way. An institution can be an actor, but an actor can also be a technology – an algorithm, for instance. Actors are actors because they do things – or the actors interact. What is of interest to social scientists is what happens when actors interact. In texts, actors in the text are also actors (e.g. politicians and countries can act). An actor is someone who has an intention, has agency. In computer science, actor means something else. One can have actor-based computations, programs that exhibit certain behavior, for instance programs that are self-regulating, self-organizing, can take in and send out messages, modeling dynamic behavior (in the macro aspect). Computer science calls actors AGENTS.

Actor network theory can be useful in looking at networks. Methods involving pattern detection, network analysis, similarity measures, time and dynamics can be used for analyzing relations/relationships in networks. Networks are useful to measure distance, so distance measures must be defined in the research design. The closeness of nodes can visually represent this. Or size can carry information. There are two levels of networks metrics: network density, how many of the possible connections exist in a network; and the structural position of a node in a network. Network analyses are node centric, focused on betweenness.

In the ‘mental model’ networks consist of several layers (questions, representations, relationships), to which is needed approximation through nodes. Distance and position are relational, while boundaries require fixity to establish a starting point to further establish centrality, density, and activity in the network. Hence, network analyses consist of metrics that quantify structural properties. Patterns in networks are communities or clusters – sub-graphs with certain properties. Patterns can be detected through clustering measures (distance for instance), identifying where clusters emerge, representing activity. Patterns can be associations that are repeating or in combination, over time, like sequences. Technical tools to collect, analyze and visualize networks include Gephi, Snappi.pi, NetworkX, and Node XL, which is an excel extension (from the work of Mark Smith). Similarity in networks indicates connection.

The group also discussed concepts like homophily, strength of weak ties, assortativity, connection principles, page rank, and bias, sampling, labeling and classification, training data, as well as ethical issues. Large part of the discussion used the problem of finding terrorists as illustrating vantage point. This angle spurred topics such as outliners and black swans, the Go Game, deep learning, the Cynefin model, gamification as incentive mechanisms to create training data, and survey design.

3.10 Workshop on What is not there yet?

Helle Sjøvaag (University of Bergen, NO)

License © Creative Commons BY 3.0 Unported license
© Helle Sjøvaag

The breakout session consisted of Ralf Schenkel (Universität Trier, DE), Damian Trilling (University of Amsterdam, NL), Tatjana Scheffler (Universität Potsdam, DE) and Helle Sjøvaag (University of Bergen, NO). The overarching questions revolved around imaginative futures for how computer science can contribute to communication science. In terms of impossible futures, the group outlined five areas for future developments:

1. Comparative, multilingual framing analysis;
2. Diverse recommender systems;
3. Automatic validation of fact statements;
4. Speech/video to text; and
5. Bot-human interaction (automated communication).

Multilingual comparative framing analysis was the most concrete envisioned future. The ‘dream scenario’ would involve a system that can map different frames (aspects of a story) of the same topic across outlets, for instance to look for diversity. This would involve computational tools that could a) identify events; b) track events; c) identify similar and diverse sources; d) group sources/themes/events into different perspective and/or link the elements; to f) predict future events. This process entails automatic translation of multiple languages, understanding different argument structures that require language independent NLP. Developments in AI/deep learning in combination with big data could serve to fulfill these dream scenarios of communication scientists.

4 Panel discussions

4.1 Data Access

Thorsten Quandt (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license
© Thorsten Quandt

The group was discussing typical problems connected to data access. We found that most of the participants were using some form of scraper, extracting (mostly) textual information from unstructured web resources. Based on the shared experiences, we tried to systematize the various access types and formats in an overview table, primarily populated by what has been done already by the workshop participants (pdf is attached to the report).

Participants

- Christian Baden
The Hebrew University of Jerusalem, IL
- David Domingo
Free University of Brussels, BE
- Martin Emmer
FU Berlin, DE
- Raul Ferrer Conill
Karlstad University, SE
- Johann-Christoph Freytag
HU Berlin, DE
- Elisabeth Günther
Universität Münster, DE
- Krishna P. Gummadi
MPI-SWS – Saarbrücken, DE
- Ari Heinonen
University of Tampere, FI
- Jukka Huhtamäki
Tampere University of Technology, FI
- Seth C. Lewis
University of Oregon, US
- Alexander Löser
Beuth Hochschule für Technik – Berlin, DE
- Wiebke Loosen
Hans-Bredow-Institut – Hamburg, DE
- Truls Pedersen
University of Bergen, NO
- Thorsten Quandt
Universität Münster, DE
- Tatjana Scheffler
Universität Potsdam, DE
- Ralf Schenkel
Universität Trier, DE
- German Shegalov
Twitter – San Francisco, US
- Helle Sjøvaag
University of Bergen, NO
- Hendrik Stange
Fraunhofer IAIS – St. Augustin, DE
- Eirik Stavelin
University of Bergen, NO
- Martin Theobald
Universität Ulm, DE
- Heike Trautmann
Universität Münster, DE
- Damian Trilling
University of Amsterdam, NL
- Gottfried Vossen
Universität Münster, DE
- Rodrigo Zamith
University of Massachusetts – Amherst, US



Report from Dagstuhl Seminar 16142

Multidisciplinary Approaches to Multivalued Data: Modeling, Visualization, Analysis

Edited by

Ingrid Hotz¹, Evren Özarslan², and Thomas Schultz³

- 1 Linköping University, SE, ingrid.hotz@liu.se
- 2 Linköping University, SE, evrenoza@gmail.com
- 3 Universität Bonn, DE, schultz@cs.uni-bonn.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16142, “Multidisciplinary Approaches to Multivalued Data: Modelling, Visualization, Analysis”, which was attended by 27 international researchers, both junior and senior. Modelling multivalued data using tensors and higher-order descriptors has become common practice in neuroscience, engineering, and medicine. Novel tools for image analysis, visualization, as well as statistical hypothesis testing and machine learning are required to extract value from such data, and can only be developed within multidisciplinary collaborations. This report gathers abstracts of the talks held by participants on recent advances and open questions related to these challenges, as well as an account of topics raised within two of the breakout sessions.

Seminar April 3–8, 2016 – <http://www.dagstuhl.de/16142>

1998 ACM Subject Classification I.4 Image Processing and Computer Vision, I.5 Pattern Recognition, J.2 Physical Sciences and Engineering, J.3 Life and Medical Sciences

Keywords and phrases visualization, image processing, statistical analysis, machine learning, tensor fields, higher-order descriptors, diffusion-weighted imaging (DWI), structural mechanics, fluid dynamics, microstructure imaging, connectomics, uncertainty visualization, feature extraction

Digital Object Identifier 10.4230/DagRep.6.4.16

Edited in cooperation with Andrada Ianaş

1 Executive Summary

Ingrid Hotz

Evren Özarslan

Thomas Schultz

License © Creative Commons BY 3.0 Unported license
© Ingrid Hotz, Thomas Schultz, and Evren Özarslan

Topics and Motivation

This seminar is the sixth in a series of Dagstuhl Seminars devoted to the use of tensor fields and other higher order descriptors, including higher-order tensors or Spherical Harmonics, to model intricate multivalued data that arises in modern medical imaging modalities, as well as in simulations in engineering and industry. Even though the literature on image analysis, visualization, as well as statistical hypothesis testing and machine learning is quite rich for scalar or vector-valued data, relatively little work has been performed on these disciplines for tensors and higher-order descriptors.



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

Multidisciplinary Approaches to Multivalued Data: Modeling, Visualization, Analysis, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 16–38

Editors: Ingrid Hotz, Evren Özarslan, and Thomas Schultz



DAGSTUHL
REPORTS Dagstuhl Reports

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

Applications wherein such descriptors can be employed to process multivalued data range from neuroimaging to image analysis and engineering. Diffusion Weighted Magnetic Resonance Imaging (DW-MRI), an MRI modality which makes it possible to visualize and quantify structural information about white matter pathways in the brain *in vivo*, is one of the driving technologies, but tensors have also shown their usefulness as feature descriptors for segmentation and grouping in image analysis, including structure tensors and tensor voting. Applications in solid mechanics, civil engineering, computational fluid dynamics and geology require the processing of tensor fields as part of domain-specific modeling, simulation, and analysis (e.g. stress-strain relationships, inertia tensors, permittivity tensor).

The Dagstuhl seminar provides a unique platform by facilitating scientific exchange between key researchers in seemingly diverse applications. Despite these disciplines' commonalities in terms of the tools employed, it would be very unlikely that these scientists would attend the same conference as the theme of most conferences is defined by a specific application. By bringing together specialists in visualization, image processing, statistics, and numerical mathematics, the Dagstuhl seminar provides new impulses for methodological work in those areas.

Organization of the Seminar

To ensure a steady inflow of new ideas and challenges, we put an emphasis on inviting researchers who previously did not have the opportunity to attend one of the meetings in this series. This was true for almost half the attendees in the final list of participants.

The seminar itself started with a round of introductions, in which all participants presented their area of work within 100 seconds with help of a single slide. This helped to create a basis for discussion early on during the week, and was particularly useful since participants came from different scientific communities, backgrounds, and countries.

A substantial part of the week was devoted to presentations by 26 participants, who spent 20 minutes each on presenting recent advances, ongoing work, or open challenges, followed by ten minutes of discussion in the plenary, as well as in-depth discussions in the breaks and over lunch. Abstracts of the presentations are collected in this report. For the traditional social event on Wednesday, we went on a hike, which was joined by almost all participants, and offered additional welcome opportunities for interaction.

Three breakout sessions were organized in the afternoons, and another one in the evening, so that none of them took place in parallel, and everyone had the opportunity to visit all groups relevant to him or her. The topics of the four groups were formed by clustering topics brought up in the round of introductions, and were denoted as:

- Visual encodings and the interface between theory and applications
- Models and geometry
- Topological methods
- Multi-field and tensor group analysis

Depending on the interests of the participants, the breakout groups differed in nature, ranging from the collection of open questions and discussions on future directions of the field to spontaneous tutorial-style presentations. Notes taken during these sessions, and the main results of two of them are summarized in this report.

Outcomes

The participants all agreed that the meeting was successful and stimulating, and we plan to publish another Springer book documenting the results of the meeting. Participants have pre-registered thirteen chapters already during the seminar, and we are in the process of collecting additional contributions both from participants and from researchers working on closely related topics who could not attend the meeting. We expect that the book will be ready for publication in 2017.

It was voted that the group will apply for another meeting in this series. In addition to the current organizers Thomas Schultz (University of Bonn, Germany) and Evren Özarlan (Linköpings Universitet, Sweden), Andrea Fuster (TU Eindhoven, The Netherlands) and Eugene Zhang (Oregon State University, USA) agreed to help apply for the next event.

Acknowledgments

The organizers thank all the attendees for their contributions and extend special thanks to the team of Schloss Dagstuhl for helping to make this seminar a success. As always, we enjoyed the warm atmosphere, which supports formal presentations as well as informal exchange of ideas.

2 Table of Contents

Executive Summary

<i>Ingrid Hotz, Thomas Schultz, and Evren Özarlan</i>	16
---	----

Overview of Talks

Composite Networks: Joint Structural-Functional Modeling of Brain <i>Burak Acar</i>	21
Interpolation of orientation distribution functions in diffusion weighted imaging <i>Maryam Afzali-Deligani</i>	21
Towards the Processing of Rotation Fields <i>Bernhard Burgeth, Andreas Kleefeld</i>	22
Geometrical Modeling in Diffusion MRI: From Riemann to Finsler <i>Tom Dela Haije, Luc Florack, and Andrea Fuster</i>	22
Bayesian heteroscedastic Rice regression for diffusion tensor imaging <i>Anders Eklund</i>	22
New possibilities with shortest-path tractography <i>Aasa Feragen</i>	23
Cartan Scalars in Finsler-DTI for Higher Order Local Brain Tissue Characterization <i>Luc Florack, Tom Dela Haije, and Andrea Fuster</i>	24
What do the Universe and the brain have in common? <i>Andrea Fuster</i>	24
Moment Invariants for multi-dimensional Data <i>Hans Hagen and Roxana Bujack</i>	25
Anisotropic Sampling for Texture Generation and Glyph Distribution <i>Ingrid Hotz</i>	25
Advanced diffusion MRI for microstructure imaging <i>Andrada Ianuş</i>	26
Tract Orientation and Angular Dispersion Deviation Indicator (TOADDI): A framework for single-subject analysis in diffusion tensor imaging <i>Cheng Guan Koay</i>	27
Magnetic Susceptibility Tensor: Imaging and Modeling <i>Chunlei Liu</i>	27
Visual Integration of Spatial-Nonspatial Data in Engineering <i>Georgeta Elisabeta Marai</i>	28
Towards the estimation of biomechanical tensors through efficient image processing methods <i>Rodrigo Moreno</i>	28
Substitutability of Symmetric Second-Order Tensor Fields: An Application in Urban 3D LiDAR Point Cloud <i>Jaya Sreevalsan Nair and Beena Kumari</i>	29
Characterizing Diffusion Anisotropy with a Confinement Tensor <i>Evren Özarlan</i>	29

Characterizing microstructural tissue abnormalities in mild traumatic brain injury with diffusion compartment imaging <i>Benoit Scherrer</i>	30
Visualization of Third-Order Tensor Fields <i>Gerik Scheuermann, Markus Stommel, Valentin Zobel</i>	30
Along-the-Tract Feature Extraction as a Manifold Learning Problem <i>Thomas Schultz</i>	31
Iteratively Reweighted L1-fitting For Model-Independent Outlier Removal And Regularization In Diffusion MRI <i>Alexandra Tobisch</i>	31
Tractography-based Edge Detection for Diffusion Weighted MRI Analysis <i>Xavier Tricoche</i>	32
Multivalued Data Processing Techniques for Diffusion MRI <i>Gözde Ünal</i>	32
Visual Group-Analysis for Diffusion Tensor Data <i>Anna Vilanova Bartroli and Changgong Zhang</i>	33
Upper bound of transition points in a 3D linear tensor field <i>Yue Zhang</i>	33
Feature-Based Visualization of Stress and Fiber Orientation Tensor Fields <i>Valentin Zobel, Gerik Scheuermann, and Markus Stommel</i>	33
Panel discussions	
Visual Encodings and Theory/Application Interface <i>Georgeta Elisabeta Marai, Maryam Afzali-Deligani, Bernhard Burgeth, Ingrid Hotz, Jaya Sreevalsan Nair, Anna Vilanova Bartroli, and Yue Zhang</i>	34
Multifield tensor analysis – group tensor analysis <i>Anna Vilanova Bartroli</i>	35
List of Previous Meetings in this Seminar Series	36
Schedule	37
Participants	38

3 Overview of Talks

3.1 Composite Networks: Joint Structural-Functional Modeling of Brain

Burak Acar (Bogaziçi University – Istanbul, TR)

License  Creative Commons BY 3.0 Unported license
© Burak Acar

Brain has long been known to be a collection of inter-connected units with a complex signaling mechanism. Our understanding of the brain has evolved from models composed of single-units responsible for tasks and their inter-connectivity towards multiple inter-connected units jointly responsible for tasks. This shift boosted research efforts on brain network analysis that has been grouped as functional and structural networks.

Functional networks (fNETs) are based on the assumption that coherence between spatially distant nodes of an fNET can be observed as the correlation between observed signals, which range from EEG, MEG to fMRI. Challenged by spatial resolution (EEG, MEG), temporal resolution (fMRI) and noise, the fNET models were provided with the substrate they needed, the structural networks (sNET), by the diffusion MRI (dMRI) which is the unique modality for in-vivo imaging of brain structure, albeit its spatial resolution limitations.

Integrating these complementary models is a major challenge towards a deeper understanding of how the brain works. While overcoming the aforementioned temporal and spatial resolution of fMRI and dMRI would be a major milestone on the road, composite network modeling requires more. The major question is whether there is a causal relationship between structure and function. This question manifests in various ways: Is there a priority / hierarchy relationship between structure and function? Who comes first? Do they change roles? Can one serve as a constraint / bias for the other? etc.

A practical approach to the big problem could be to pursue a clinical perspective. In other words, it seems promising to ask how one can answer the clinical questions best by considering the structure and function simultaneously. Our on-going multi-center BRAINetc project pursues this path for the Alzheimer's Disease (AD).

3.2 Interpolation of orientation distribution functions in diffusion weighted imaging

Maryam Afzali-Deligani (Sharif University of Technology – Tehran, IR)

License  Creative Commons BY 3.0 Unported license
© Maryam Afzali-Deligani

Diffusion weighted imaging (DWI) is a non-invasive method for investigating the brain white matter structure and can be used to evaluate fiber bundles. However, due to practical constraints, DWI data acquired in clinics are low resolution. We propose a method for interpolation of orientation distribution functions (ODFs). To this end, fuzzy clustering is applied to segment ODFs based on the principal diffusion directions (PDDs). Next, a cluster is modeled by a tensor so that an ODF is represented by a mixture of tensors. For interpolation, each tensor is rotated separately. The proposed method is appropriate for increasing resolution in the ODF field and can be applied to clinical data to improve evaluation of white matter fibers in the brain.

3.3 Towards the Processing of Rotation Fields

Bernhard Burgeth (Universität des Saarlandes, DE), Andreas Kleefeld (Universität des Saarlandes, DE)

License  Creative Commons BY 3.0 Unported license
© Bernhard Burgeth, Andreas Kleefeld

A rotation field can be considered as a mapping from an image domain into the set of orthonormal $n \times n$ -matrices, $n = 2, 3, \dots$ with determinant $+1$. Rotation fields might originate from fields of symmetric matrices via eigendecomposition and as such potentially play a role, for example, in medical imaging and material science. Although orthonormal matrices have many algebraic properties it is not clear how to extend image processing methods to this type of data. Even standard interpolation of orthonormal matrices is not straight forward. In this talk we present approaches to create building blocks (averaging, supremum, infimum etc.) for elementary image processing of rotation fields. After presenting some preliminary results we will discuss prerequisites, shortcomings, and potential of the proposed concepts.

3.4 Geometrical Modeling in Diffusion MRI: From Riemann to Finsler

Tom Dela Haije (TU Eindhoven, NL), Luc Florack (TU Eindhoven, NL), and Andrea Fuster (TU Eindhoven, NL)

License  Creative Commons BY 3.0 Unported license
© Tom Dela Haije, Luc Florack, and Andrea Fuster

In geometrical modeling of diffusion MRI, spin dynamics in space are assumed to correspond to a simple stochastic (Brownian) process on a manifold. In this work we introduce the basic concepts of two geometrical models used in diffusion MRI, in which the tissue is modeled as either a Riemannian or a Finslerian manifold. The Riemannian framework for diffusion MRI was originally proposed in 2002, and since then considerable effort has been made to extend it to the more complex Finslerian case. We recently introduced a canonical definition for the Finslerian geometrical structure in terms of the diffusion MRI signal, which solves one of the major hurdles discussed at the last Dagstuhl meeting.

3.5 Bayesian heteroscedastic Rice regression for diffusion tensor imaging

Anders Eklund (Linköping University, SE)

License  Creative Commons BY 3.0 Unported license
© Anders Eklund

Joint work of Anders Eklund, Bertil Wegmann, Mattias Villani

Diffusion weighted imaging (DWI) has during the last years been improved using better MR scanners and more advanced gradient sequences. In this presentation, the focus will instead be on improving the statistical analysis. For example, to estimate a diffusion tensor in each voxel, a standard approach is to take the logarithm of the measurements and then calculate the best parameters using least squares or weighted least squares. Additionally, the standard approaches only return a point estimate of the tensor, and ignore the uncertainty of the

estimates (which may be important for a group analysis). A more proper way to estimate the tensor parameters is to use a generalized linear model with a logarithmic link function. Using a Bayesian framework, a (posterior) distribution of the tensor parameters will be achieved, instead of a point estimate. Specifically, the use of a generalized linear model with support for heteroscedastic (non-stationary variance) Rician noise will be discussed in this presentation.

3.6 New possibilities with shortest-path tractography

Aasa Feragen (*University of Copenhagen, DK*)

License  Creative Commons BY 3.0 Unported license
© Aasa Feragen

Joint work of Niklas Kasenburg, Michael Schober, Matthew Liptrot, Nina Reislev, Silas Ørting, Mads Nielsen, Ellen Garde, Philipp Hennig, Søren Hauberg, Aasa Feragen

Main reference N. Kasenburg, M. Liptrot, N. Linde Reislev, S. N. Ørting, M. Nielsen, E. Garde, A. Feragen, “Training shortest-path tractography: Automatic learning of spatial priors”, *NeuroImage*, Vol. 130, pp. 63–76, 2016.

URL <http://dx.doi.org/10.1016/j.neuroimage.2016.01.031>

Main reference S. Hauberg, M. Schober, M. Liptrot, P. Hennig, A. Feragen, “A Random Riemannian Metric for Probabilistic Shortest-Path Tractography”, in *Proc. of the 18th Int’l Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI’15)*, LNCS, Vol. 9349, pp. 597–604, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-24553-9_73

Tractography is a family of algorithms that aim to estimate the trajectories of brain fibers from noisy diffusion weighted MRI data. The most typical approach for estimating such trajectories is *fiber tracking*, where the algorithm starts at a pre-selected seed point and keeps walking in an estimated “most likely” direction until a stopping criterion is reached. Tracking methods suffer from “path length dependency” which results in a) propagation of uncertainty with distance to the seed point, and b) decrease in the probability of ever reaching a target point with its distance to the seed point, regardless of whether there is a physical connection or not.

An alternative approach is *shortest-path tractography* (SPT), which reformulates tractography as a shortest path problem either on a graph or in a Riemannian manifold. I will discuss two new tools made possible by shortest-path tractography:

- Learning spatial priors for graph-based SPT, allowing learned or existing prior knowledge to improve tractography output.
- Representing tractography output as a Gaussian Process probability distribution over curves in a Riemannian manifold, allowing uncertainty estimates that do not propagate with seed point distance.

3.7 Cartan Scalars in Finsler-DTI for Higher Order Local Brain Tissue Characterization

Luc Florack (TU Eindhoven, NL), Tom Dela Haije (TU Eindhoven, NL), and Andrea Fuster (TU Eindhoven, NL)

License  Creative Commons BY 3.0 Unported license
© Luc Florack, Tom Dela Haije, and Andrea Fuster

Joint work of Luc Florack, Andrea Fuster, Tom Dela Haije

Main reference L. M. J. Florack, T. C. J. Dela Haije, A. Fuster, “Direction-controlled DTI interpolation”, in I. Hotz, T. Schultz (eds.), “Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data”, Mathematics and Visualization, pp. 149–162, Springer, 2015.

URL <http://dx.doi.org/10.1007/978-3-319-15090-1>

Main reference L. M. J. Florack, A. Fuster, “Riemann-Finsler geometry for diffusion weighted magnetic resonance imaging”, in C. F. Westin, A. Vilanova, B. Burgeth (eds.), “Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data”, Mathematics and Visualization, pp. 189–208, Springer, 2014.

URL <http://dx.doi.org/10.1007/978-3-642-54301-2>

In diffusion weighted magnetic resonance imaging (dwMRI) there is a need for “higher order” data representations. Geometric representations are of particular interest. In previous work we advocated the use of Finsler geometry to generalize the Riemannian framework developed in the context of diffusion tensor imaging (DTI). The latter stipulates that diffusivity along a path can be quantified in terms of a data-adapted “length” functional, for which the (inverse) DTI tensor provides the defining Gram matrix for an inner product norm (6 d.o.f.’s per spatial position). The Finsler-DTI extension likewise stipulates a data-adapted length functional in terms of a second order tensor. The corresponding norm is given by a generalized “Finsler-DTI” tensor that lives on a 5-dimensional manifold of space and orientation (formally 6 d.o.f.’s, but effectively reducible to a single d.o.f. per spatial position and orientation). Unlike with DTI, this Finsler norm is not (necessarily) induced by an inner product, and admits an unlimited number of local d.o.f.’s (with each orientation treated as an a priori independent variable). The (third order) Cartan tensor of Finsler geometry captures the residual d.o.f.’s discarded in the classical Riemann-DTI rationale. It is of interest to study scalars induced by this tensor, as they disclose features of local fiber architecture that DTI fails to capture. As such they complement well-known ones from classical DTI, such as fractional isotropy, and are amenable to traditional image analysis techniques for classification, visualization, etc.

3.8 What do the Universe and the brain have in common?

Andrea Fuster (TU Eindhoven, NL)

License  Creative Commons BY 3.0 Unported license
© Andrea Fuster

In this talk I investigate similarities between techniques used in the analysis and visualization of astronomy and brain imaging data.

3.9 Moment Invariants for multi-dimensional Data

Hans Hagen (TU Kaiserslautern, DE) and Roxana Bujack (TU Kaiserslautern, DE)

License © Creative Commons BY 3.0 Unported license
© Hans Hagen and Roxana Bujack

Moment invariants have long been successfully used for pattern matching in scalar fields. By their means, features can be detected in a data set independent of their exact orientation, position, and scale. Their recent extension to vector fields was the first step towards rotation invariant pattern detection in multi-dimensional data. We research the state of the art on moment invariants for vector valued data and evaluate the potential of the different approaches for the next step of generalizing moment invariants to tensor fields.

3.10 Anisotropic Sampling for Texture Generation and Glyph Distribution

Ingrid Hotz (Linköping University, SE)

License © Creative Commons BY 3.0 Unported license
© Ingrid Hotz

Joint work of Andrea Kratz, Ingrid Hotz

Main reference A. Kratz, D. Baum, I. Hotz, “Anisotropic Sampling of Planar and Two-Manifold Domains for Texture Generation and Glyph Distribution”, *Transaction on Visualization and Computer Graphics*, 19(11):1782–1794, IEEE, 2013.

URL <http://dx.doi.org/10.1109/TVCG.2013.83>

Anisotropic sample distributions on planar and two-manifold domains are useful for many applications. Our work has been motivated by generating uniform but still aperiodic glyph distributions and texture generation. The requirement for the sampling is being dense, covering the entire surface and being aperiodic to not generating artificial visual patterns. A second requirement for our work is efficiency and robustness. The sample generation should be interactive and work without tedious parameter tuning even for rapidly changing size and orientation of the samples. To reach this goal we employ an anisotropic triangulation that serves as basis for the creation of an initial sample distribution as well as for a gravitational-centered relaxation. Furthermore, define anisotropic Voronoi cells as base element for texture generation. It represents a novel and flexible visualization approach to depict metric tensor fields that can be derived from general tensor fields as well as scalar or vector fields.

3.11 Advanced diffusion MRI for microstructure imaging

Andrada Ianuş (University College London, GB)

License © Creative Commons BY 3.0 Unported license
© Andrada Ianuş

Joint work of Andrada Ianuş, Ivana Drobnjak, Gary Hui Zhang, Enrico Kaden, Daniel C. Alexander
Main reference I. Drobnjak, H. Zhang, A. Ianuş, E. Kaden, D. C. Alexander, “PGSE, OGSE, and sensitivity to axon diameter in diffusion MRI: Insight from a simulation study”, *Magn Reson Med.*, 75(2):688–700, 2016.

URL <http://dx.doi.org/10.1002/mrm.25631>

Main reference A. Ianuş, I. Drobnjak, D. C. Alexander, “Model-based estimation of microscopic anisotropy using diffusion MRI: a simulation study”, *NMR in Biomedicine*, 29(5):672–685, 2016.

URL <http://dx.doi.org/10.1002/nbm.3496>

Diffusion weighted MRI (DW-MRI) probes the displacement of water molecules inside the tissue, which is influenced by the presence of cellular membranes. Microstructure imaging techniques use mathematical models which describe the effect of various tissue properties (e.g. cellular size, shape, volume fraction, etc) on the acquired signal and fits them to the data, in order to infer microscopic features from images that have a much lower resolution (usually at the millimetre scale). Here I present some of the current and future directions of microstructure imaging techniques developed in our group, with applications to both brain and cancer imaging.

The first part is focused on applications of microstructure imaging in the brain. Measuring axon diameter provides potential biomarkers for staging and monitoring the progression of white matter diseases such as multiple sclerosis, as well as for understanding the ageing process. The current techniques for mapping axon diameter, such as AxCaliber or Active Ax, use a collection of standard single diffusion encoding (SDE) sequences, which do not provide the optimal sensitivity to axon diameter. We have compared in simulation the sensitivity of SDE and oscillating diffusion encoding (ODE) sequences to axon diameter, showing that in practical situations of dispersed axons, as well as multiple gradient orientations, ODE sequences are beneficial for estimating axon diameter. We also discuss the resolution limit of this technique, i.e. the smallest diameter that can be distinguished.

The second part is focused on diffusion acquisition and modelling techniques with potential applications for cancer imaging. Cellular anisotropy is an important microstructural feature, which has the potential to distinguish between different types of tumours as well as different tumour grades. A recently developed technique for cancer imaging, namely VERDICT MRI, which uses a collection of SDE sequences and models multi-compartment diffusion, does not account for pore eccentricity. A widely used sequence in the literature for estimating microscopic anisotropy is double diffusion encoding, which vary the gradient orientation within one measurements, however, most of studies do not recover intrinsic estimates of pore size and eccentricity. Here I present a model-based approach that allows the estimation of pore size and eccentricity in complex substrates which consist of elongated pores with a distribution of sizes, as well as a comparison between the ability of SDE and DDE sequences to recover the ground truth microstructural parameters depending on the complexity of the substrates.

3.12 Tract Orientation and Angular Dispersion Deviation Indicator (TOADDI): A framework for single-subject analysis in diffusion tensor imaging

Cheng Guan Koay (Walter Reed Medical Center – Bethesda, US)

License © Creative Commons BY 3.0 Unported license
© Cheng Guan Koay

The purpose of the proposed framework is to carry out single-subject analysis of diffusion tensor imaging (DTI) data. This framework is termed Tract Orientation and Angular Dispersion Deviation Indicator (TOADDI). It is capable of testing whether an individual tract as represented by the major eigenvector of the diffusion tensor and its corresponding angular dispersion are significantly different from a group of tracts on a voxel-by-voxel basis. This work develops two complementary statistical tests (orientation and shape tests) based on the elliptical cone of uncertainty, which is a model of uncertainty or dispersion of the major eigenvector of the diffusion tensor.

3.13 Magnetic Susceptibility Tensor: Imaging and Modeling

Chunlei Liu (University of California – Berkeley, US)

License © Creative Commons BY 3.0 Unported license
© Chunlei Liu

Magnetic susceptibility is a quantitative measure of the extent to which a material is magnetized by an applied magnetic field. The magnetic susceptibility of a material, noted by χ , is equal to the ratio of the magnetization M within the material to the applied magnetic field strength H , i.e. $\chi = M/H$. Susceptibility anisotropy describes that magnetic susceptibility is a tensor quantity rather than a scalar quantity. Susceptibility tensor imaging (STI) was proposed to measure and quantify susceptibility as a rank-2 tensor. This technique relies on the measurement of frequency offsets at different orientations with respect to the main magnetic field. The orientation dependence of susceptibility is characterized by a tensor. In the brain's frame of reference, the relationship between frequency shift and susceptibility tensor is given by:

$$f(\mathbf{k}) = \gamma B_0 \left(\frac{1}{3} \hat{\mathbf{H}}^T \chi(\mathbf{k}) \hat{\mathbf{H}} - \hat{\mathbf{H}} \cdot \mathbf{k} \frac{\mathbf{k}^T \chi(\mathbf{k}) \hat{\mathbf{H}}}{k^2} \right) \quad (1)$$

Here, χ is a second-order (or rank-2) susceptibility tensor; $\hat{\mathbf{H}}$ is the unit vector (unitless) of the applied magnetic field. Assuming that the susceptibility tensor is symmetric, then there are six independent variables to be determined for each tensor. In principle, a minimum of six independent measurements are necessary. A set of independent measurements can be obtained by rotating the imaging object, e.g. tilting the head, with respect to the main magnetic field. Given a set of such measurements, a susceptibility tensor can be estimated by inverting the system of linear equations formed by Eq. 1. Fewer than six orientations are also feasible by incorporating fiber orientation estimated by diffusion tensor imaging (DTI) and assuming cylindrical symmetry of the susceptibility tensor. Susceptibility tensor can be decomposed into three eigenvalues (principal susceptibilities) and associated eigenvectors. Similar to DTI fiber tractography, fiber tracts can be reconstructed based on STI.

3.14 Visual Integration of Spatial-Nonspatial Data in Engineering

Georgeta Elisabeta Marai (University of Illinois – Chicago, US)

License © Creative Commons BY 3.0 Unported license
© Georgeta Elisabeta Marai

Joint work of Fillipo Pellolio, Chihua Ma, Timothy Luciani, Georgeta Elisabeta Marai

Multifield engineering data features sometimes both spatial and nonspatial characteristics. In the visualization field, “spatial data” denotes datasets whose input attributes specify the position of each item – for example airflow around an airplane wing. In contrast, “nonspatial data” denotes completely abstract datasets in which no attributes have intrinsic spatial position semantics—for example sets or tables. In this context, are tensors spatial or nonspatial quantities? Does the distinction matter? What can we learn from the existing visual integration designs which exist for engineering data?

3.15 Towards the estimation of biomechanical tensors through efficient image processing methods

Rodrigo Moreno (KTH Royal Institute of Technology – Stockholm, SE)

License © Creative Commons BY 3.0 Unported license
© Rodrigo Moreno

Joint work of Örjan Smedby, Dieter Pahr, Rodrigo Moreno

Main reference R. Moreno, Ö. Smedby, D.H. Pahr, “Prediction of apparent trabecular bone stiffness through fourth-order fabric tensors”, *Biomechanics and Modeling in Mechanobiology*, 15(4):831–844, Springer, 2015.

URL <http://dx.doi.org/10.1007/s10237-015-0726-5>

Tensors are widely used in biomechanics to describe different physical properties of tissue. Although diffusion MRI has been very successful describing anisotropy and orientation of tissue non invasively, its use is mainly restricted to certain anatomical sites, such as the brain and muscle fibers. In other applications, tensors can be obtained either through physical experiments or simulations using anatomical images as an input. Both options are inconvenient in clinics as experiments are usually not possible and simulations are time consuming. Alternatively, a good and efficient approximation of these tensors can be obtained by finding connections between geometry descriptors of the tissue and the tensorial variables of interest. In this talk, I will show the use of this approach in two different contexts. In the first one, the stiffness tensor, which is one of the most important biomechanical parameter of trabecular bone, is estimated using fabric tensors [1]. In the second one, the permeability tensor of the microvasculature of the liver, which is important to understand the micro perfusion process in the liver, is also approximated through fabric tensors. In these two applications, the estimations are obtained in just a few seconds with a relatively high accuracy, compared to the very expensive FEM-based approaches. The preliminary results suggest that geometry plays a big role in these physical entities. Our current efforts aim at finding further connections between more advanced geometrical descriptors and different biomechanical tensors.

3.16 Substitutability of Symmetric Second-Order Tensor Fields: An Application in Urban 3D LiDAR Point Cloud

Jaya Sreevalsan Nair (IIIT – Karnataka, IN) and Beena Kumari

License © Creative Commons BY 3.0 Unported license

© Jaya Sreevalsan Nair and Beena Kumari

Joint work of Beena Kumari, Jaya Sreevalsan-Nair

Main reference B. Kumari, J. Sreevalsan-Nair, “An interactive visual analytic tool for semantic classification of 3D urban LiDAR point cloud”, in Proc. of the 23rd SIGSPATIAL Int’l Conf. on Advances in Geographic Information Systems (GIS’15), Article No. 73, ACM, 2015.

URL <http://dx.doi.org/10.1145/2820783.2820863>

There has been work on augmented semantic classification of 3D urban LiDAR point cloud, which is a tuple of class labels from two different classifications, namely structural and contextual. The goal of augmented semantic classification is to extract curves (boundaries, ridges, etc.) and identify objects (buildings, vegetation, asphalt and natural ground), which will further enable 3D object extraction. The structural classification, which is an essential step in augmented semantic classification, is computed using a multi-scale approach of using structure tensor, which effectively defines di Zenzo multi-valued geometry. Spectral values of the structure tensor have been used frequently in LiDAR community to derive multiple values of local geometry. While the structure tensor, computed probabilistically using covariance matrix, encodes proximity and continuity, we propose the use of an anisotropic diffusion tensor, derived from voting tensor, to encode information on weights on proximity and the global context of the local neighborhood. Our proposed voting tensor additionally uses eigenvector orientation for determining diffusion velocity. The goal of our work is to improve the outcomes of structural classification and the overall augmented classification, where we use unsupervised methods. We further discuss the extent of substitutability of the tensor field, constructed from covariance matrix, with a diffused normal voting tensor field – in addition to the application of classification where the substitution works, we show an application in local geometry extraction where the substitution does not work.

3.17 Characterizing Diffusion Anisotropy with a Confinement Tensor

Evren Özarslan (Linköping University, SE)

License © Creative Commons BY 3.0 Unported license

© Evren Özarslan

We studied the influence of diffusion on NMR experiments when the diffusing molecules are subjected to a force field. We place special emphasis on parabolic (Hookean) potentials, which we tackled theoretically using path integral methods. We obtained explicit relationships for commonly employed gradient waveforms involving pulsed and oscillating gradients. Semi-analytical multiple correlation function (MCF) method as well as random walk simulations validated our theoretical results.

The three-dimensional formulation of the problem leads to a new characterization of diffusional anisotropy. Unlike for the case of traditional methods that employ a diffusion tensor, anisotropy in our model originates from the stiffness tensor of a virtual spring while bulk diffusivity is retained in the formulation. Our approach thus yields an expansive alternative to diffusion tensor imaging (DTI). Contrary to DTI, our technique accounts for the restricted character of the diffusion process as reflected in its diffusion-time dependence. The

formalism is expected to be useful in addressing a variety of problems involving macroscopic (global) and microscopic (local) diffusion anisotropy.

3.18 Characterizing microstructural tissue abnormalities in mild traumatic brain injury with diffusion compartment imaging

Benoit Scherrer (Harvard Medical School – Boston, US)

License © Creative Commons BY 3.0 Unported license
© Benoit Scherrer

Main reference B. Scherrer, A. Schwartzman, M. Taquet, M. Sahin, S. P. Prabhu, S. K. Warfield, “Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in diffusion-compartment imaging (DIAMOND)”, in *Magn Reson Med*, 2015, to appear.

While diffusion tensor imaging (DTI) has proven sensitive to detecting microstructural changes in mild traumatic injury (mTBI), it lacks specificity and fails at providing a mechanistic insight into tissue changes. mTBI is associated with cytotoxic edema, neuroinflammation and traumatic axonal injury (TAI) of varying severity, each of which have a unique intra-voxel diffusion signature when imaging with multiple b-values. Diffusion compartment imaging (DCI) aims at teasing apart the distinct types of diffusion within voxels whenever they are present, providing improved insight into underlying tissues changes. We present recent developments in DCI and provide preliminary evidence that DCI enables more detailed characterization of tissue changes in mTBI with higher sensitivity and specificity.

3.19 Visualization of Third-Order Tensor Fields

Gerik Scheuermann (Universität Leipzig, DE), Markus Stommel (TU Dortmund, DE) and Valentin Zobel (Universität Leipzig, DE)

License © Creative Commons BY 3.0 Unported license
© Gerik Scheuermann, Markus Stommel, Valentin Zobel

In some applications, it is necessary to look into gradients of (symmetric) second order tensor fields. These tensors are of third order. In three-dimensional space, we have 18 independent coefficients at each position, so the visualization of these fields provides a challenge. A particular case are stress gradients in structural mechanics. We present specific situations where the stress gradient is required together with the stress to study material behavior. Since the visualization community lacks methods to show these fields, we look at some preliminary ideas to design appropriate glyphs. We motivate our glyph designs by typical depictions of stress in engineering textbooks.

3.20 Along-the-Tract Feature Extraction as a Manifold Learning Problem

Thomas Schultz (Universität Bonn, DE)

License © Creative Commons BY 3.0 Unported license
© Thomas Schultz

Joint work of Mohammad Khatami, Tobias Schmidt-Wilcke, Pia C. Sundgren, Amin Abbasloo, Bernhard Schölkopf, Thomas Schultz

Main reference M. Khatami, T. Schmidt-Wilcke, P. C. Sundgren, A. Abbasloo, B. Schölkopf, T. Schultz, “BundleMAP: Anatomically Localized Features from dMRI for Detection of Disease”, in Proc. of the Int’l Workshop on Machine Learning in Medical Imaging (MLMI), LNCS, Vol. 9352, pp. 52–60, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-24888-2_7

Supervised classification and regression based on diffusion MR data require the definition of suitable feature vectors, which are most frequently derived either from individual voxels, or from tractography-based networks. We demonstrate the benefits of an intermediate approach, which aggregates information along fiber bundles, while still preserving the ability to localize the effects of disease. This leads us to revisit the classical problem of along-the-tract analysis, for which we propose a novel, simple yet reliable and stable approach: We consider the joint parametrization of fiber bundles as a problem of mapping to a latent fiber core manifold, which is solved in a simple and stable manner using ISOMAP. We present BundleMAP, an integrated machine learning framework that uses this idea for classification, regression, statistical analysis and visualization, and we discuss its future integration into a multimodal framework.

3.21 Iteratively Reweighted L1-fitting For Model-Independent Outlier Removal And Regularization In Diffusion MRI

Alexandra Tobisch (DZNE – Bonn, DE & Universität Bonn, DE)

License © Creative Commons BY 3.0 Unported license
© Alexandra Tobisch

Joint work of Alexandra Tobisch, Tony Stöcker, Samuel Groeschel, Thomas Schultz

Main reference A. Tobisch, T. Stöcker, S. Groeschel, T. Schultz, “Iteratively Reweighted L1-fitting For Model-Independent Outlier Removal And Regularization In Diffusion MRI”, in Proc. of the IEEE Int’l Symp. on Biomedical Imaging (ISBI’16), pp. 911–914, IEEE, 2016.

URL <http://dx.doi.org/10.1109/ISBI.2016.7493413>

Diffusion MRI provides the possibility to investigate the structural connectivity of brain white matter non-invasively and to examine pathological conditions of the central nervous system. However, the technique is negatively affected by subject motion occurring during the image acquisition. Spatially and temporally varying artifacts, e.g. induced by subject motion, potentially degrade the signal quality and adversely influence the estimation of microstructural diffusion measures. Especially in clinical applications or large population studies, when data is collected from diseased patients, children or elderly people, measures need to be taken against the image degradation due to frequently occurring motion artifacts. State-of-the-art procedures for outlier removal detect and reject defective images during model fitting. These methods, however, are tailored only for specific diffusion models and excluding a varying number of diffusion-weighted images might be disadvantageous for the parameter estimation. We present a novel method based on an iteratively reweighted L1-Fitting for model-independent outlier removal with subsequent reconstruction of the full set of DWIs from the sparse set of inliers by modeling the signal in the continuous SHORE basis. Our

results on simulation data and clinical in vivo human brain scans demonstrate that this method corrects dMRI data for motion artifacts and reduces the impact of defective DWIs on diffusion measures.

3.22 Tractography-based Edge Detection for Diffusion Weighted MRI Analysis

Xavier Tricoche (Purdue University – West Lafayette, US)

License © Creative Commons BY 3.0 Unported license
© Xavier Tricoche

Joint work of Ziang Ding, Yaniv Gur, Xavier Tricoche

We present a technique to automatically characterize the geometry of important anatomical structures in diffusion weighted MRI (DWI) data. Our approach is based on the interpretation of diffusion data as a superimposition of multiple line fields that each form a continuum of space filling curves. Using a dense tractography computation, we quantify the spatial variations of the geometry of these curves and use the resulting measure to characterize salient structures as edges. Anatomically, these structures have a boundary-like nature and yield a clear and precise picture of major fiber bundles. Our framework leverages high angular resolution imaging (HARDI) data to offer a precise geometric description of subtle anatomical configurations associated with the local presence of multiple fiber orientations. We evaluate our technique and study its robustness to noise in the context of a phantom dataset and present results obtained with in vivo human and small animal imaging.

3.23 Multivalued Data Processing Techniques for Diffusion MRI

Gözde Ünal (Istanbul Technical University, TR)

License © Creative Commons BY 3.0 Unported license
© Gözde Ünal

This work presents two novel techniques for processing of MRI data in order to extract structural asymmetries. First technique consists of an effective regularization technique for capturing inherent asymmetry of the underlying intravoxel geometry that exists in bending, crossing or kissing fibers of the brain white matter. This, to our knowledge, is the first study that demonstrates the asymmetry at the voxel level. The second technique uses higher order tensors in modelling of tree-like structures such as vascular trees in human brain. We show how we embed the tensor in a 4D space rather than 3D in order to untangle the bifurcating (or even n-furcating) structures/branches in the data in a higher-dimensional space.

3.24 Visual Group-Analysis for Diffusion Tensor Data

Anna Vilanova Bartroli (TU Delft, NL) and Changgong Zhang (TU Delft, NL)

License © Creative Commons BY 3.0 Unported license

© Anna Vilanova Bartroli and Changgong Zhang

Joint work of Changgong Zhang, Thomas Schultz, Kai Lawonn, Elmar Eisemann, Anna Vilanova Bartroli,

Main reference C. Zhang, T. Schultz, K. Lawonn, E. Eisemann, A. Vilanova, “Glyph-Based Comparative Visualization of Diffusion Tensor Fields”, *Trans. Vis. Comput. Graph.*, 22(1):797–806, IEEE, 2016.

URL <http://dx.doi.org/10.1109/TVCG.2015.2467435>

For several applications it is necessary to compare diffusion tensor fields, e.g., to study the effects of acquisition parameters, or to investigate the influence of pathology on white matter structures. This comparison is commonly done by extracting scalar information out of the tensor fields and then comparing these scalar fields, which leads to a loss of information. If the local full tensor representation, i.e., glyphs is kept representation simple juxtaposition or superposition exist and is used, but they are not efficient visual encodings to identify difference. We propose the Tender Glyph, a glyph that explicitly encodes differences between orientation, shape and scale. However, this is limited to two tensors. Can we extend this concept to the analysis of groups of tensors? What is the box plot diagram of a group of diffusion tensors?

3.25 Upper bound of transition points in a 3D linear tensor field

Yue Zhang (Oregon State University, US)

License © Creative Commons BY 3.0 Unported license

© Yue Zhang

Transition points are an integral part of the 3D tensor field topology. They are the degenerate points separating wedges and trisectors along a degenerate curve. Much study has focused on locating degenerate points and less focus has been given to automatically extracting transition points. In this research, we provide an analytic formulation that leads to a theoretical upper bound for the number of transition points for any give 3D linear tensor field.

3.26 Feature-Based Visualization of Stress and Fiber Orientation Tensor Fields

Valentin Zobel (Universität Leipzig, DE), Gerik Scheuermann (Universität Leipzig, DE), and Markus Stommel

License © Creative Commons BY 3.0 Unported license

© Valentin Zobel, Gerik Scheuermann, and Markus Stommel

Main reference V. Zobel, M. Stommel, G. Scheuermann, “Feature-based tensor field visualization for fiber reinforced polymers”, in the Proc. of the 2015 IEEE Scientific Visualization Conference (SciVis’15), pp. 49–56, IEEE, 2015.

URL <http://dx.doi.org/10.1109/SciVis.2015.7429491>

The failure of components made from fiber reinforced polymers depends not only on the stress, it depends also on the fiber orientation. The stress is given by a stress tensor field, the fiber orientation by a fiber orientation tensor field. Both tensor fields have to be considered for the prediction of failure. In our work, we define features which indicate failure for a given

load condition in dependence of the fiber orientation. Moreover, we use glyphs to show the given and the desired fiber orientation. Since the fiber orientation can be influenced by the production process, these visualizations help the engineer to obtain fiber orientations which lead to more stable components.

4 Panel discussions

4.1 Visual Encodings and Theory/Application Interface

Georgeta Elisabeta Marai (University of Illinois – Chicago, US), Maryam Afzali-Deligani (Sharif University of Technology – Tehran, IR), Bernhard Burgeth (Universität des Saarlandes, DE), Ingrid Hotz (Linköping University, SE), Jaya Sreevalsan Nair (IIT – Karnataka, IN), Anna Vilanova Bartroli (TU Delft, NL), and Yue Zhang (Oregon State University, US)

License © Creative Commons BY 3.0 Unported license
 © Georgeta Elisabeta Marai, Maryam Afzali-Deligani, Bernhard Burgeth, Ingrid Hotz, Jaya Sreevalsan Nair, Anna Vilanova Bartroli, and Yue Zhang

The panel discussion was motivated by the challenge to design the analysis and visualization tasks in accordance with the application needs. How can analysis results be communicated to the application scientist despite the complexity of the data and the diversity of the applications. The discussion started with two major questions:

1. Which visual encodings should we use for tensor visualization?
2. How application-specific must these encodings be?

The result of a very vivid and controversial discussion was the agreement that the basis for the development of meaningful analysis and visual representations is a proper classification maybe resulting in a kind of ‘tensor catalog’ in terms of

- applications with their specific needs, questions and problems,
- tensor type from of mathematical as well as a semantic point of view,
- specific data related challenges.

Especially for the work within the field of tensor field visualization the idea of ‘one-size fits all’ is problematic. It was recognized that there is some useful work dealing with this task however it is still far from being complete. Examples that have been discussed were the classification as part of a state of the art report “Visualization and Analysis of Second-Order Tensors: Moving Beyond the Symmetric Positive-Definite Case” by Kratz et al. or the introduction of diverse parametrizations of the space of tensor fields as in “Asymmetric tensor analysis for flow visualization” by Zhang et al. or “Orthogonal Tensor Invariants and the Analysis of Diffusion Tensor Magnetic Resonance Images” by Ennis et al.

During the break-out session, we started with a collection of keywords which might be relevant for the classification which is stated below. However, due to the limited time, it did not get beyond the current state of the art. However, it demonstrated clearly that there is the the need to approach this task in a more concise way.

Keyword collection:

1. Type and properties: symmetry, definiteness, sparseness, invertibility, co- and contra-variance
2. Problem characteristics: scale of data sets, dimension (2d/3d/Nd,time-dependent), grid structure
3. Features of interest: principal directions (eigenvectors), rotational features, topology, discontinuity

4. Visualizations: level of detail, summarization, statistics, clustering
5. The encodings: colors, Glyphs, tensorlines, Topology, Volume Rendering

4.2 Multifield tensor analysis – group tensor analysis

Anna Vilanova Bartroli (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Anna Vilanova Bartroli

In this break out session, we discussed the analysis of groups of tensors. The following notes have been taken:

- Analysis of group of tensors field is interesting for several reasons: Comparison between two fields, Uncertainty description, population description, comparison between populations We discussed mathematical/statistical frameworks that exist to analyze this data
- Data tensors might be of interest to be used in this context. They can be used to summarize a whole population. Methods like alternating least squares. How do you preserve the tensor structure in such a context? This was unclear for some of the participants. Relevant literature includes:
 - Tensor Decompositions and Applications, Tamara G. Kolda, and Brett W. Bader, SIAM review
 - Renato Pajarola has also done work in this direction in the visualization community
- There are methods in the literature to calculate means and tensor covariance matrices. We discussed some of these approaches. Usually, they are coupled to different distance metrics for tensors that have been defined in the past. Relevant literature:
 - Spectral decomposition of a 4th-order covariance tensor: Applications to diffusion tensor MRI – Peter J. Basser and Sinisa Pajevic
 - Statistics on the Manifold of Multivariate Normal Distributions: Theory and Application to Diffusion Tensor MRI Processing – Christophe Lenglet and Mikaël Rousson
- Mathematical framework for multi-tensor models: Linear combination of multi-tensor models for registration, atlas and detection of statistically significant differences along fascicles using null hypothesis testing (using clustered-based statistics)
 - A Mathematical Framework for the Registration and Analysis of Multi-Fascicle Models for Population Studies of the Brain Microstructure, Maxime Taquet, Benoit Scherrer, Olivier Commowick, Jurriaan Peters, Mustafa Sahin, Benoit Macq and Simon K. Warfield, IEEE Transactions on Medical Imaging, 2014 http://perso.uclouvain.be/maxime.taquet/documents/taquet_tmi2013.pdf
 - A Framework for the Analysis of Diffusion Compartment Imaging (DCI), Maxime Taquet, Benoit Scherrer and Simon K. Warfield, in Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data, Springer 2015. http://perso.uclouvain.be/maxime.taquet/documents/taquet_springer2014.pdf
- Improvements with a new model that better preserves microstructural features (i.e., “more linear” interpolation of microstructural parameters, less swelling)
- Work on Minkowski tensors with applications in physical applications: <http://csgb.dk/research-topics/wp1/>
- There is not much visualization research in this direction, to our knowledge, the papers that are there in this direction are:

- Glyph-based Comparative Visualization for Diffusion Tensor Fields Changgong Zhang, Thomas Schultz, Kai Lawonn, Elmar Eisemann, and Anna Vilanova In: IEEE Trans. on Visualization and Computer Graphics (2016), 22:1
- Visualizing Tensor Normal Distributions at Multiple Levels of Detail Amin Abbasloo, Vitalis Wiens, Max Hermann, and Thomas Schultz In: IEEE Trans. on Visualization and Computer Graphics (2016), 22:1

5 List of Previous Meetings in this Seminar Series

- The initial Dagstuhl Perspective Workshop (Seminar 04172, April 2004, Organizers: Hans Hagen and Joachim Weickert) was the first international forum where leading experts on visualization and processing of tensor fields had the opportunity to meet, many for the first time. This workshop identified several key issues and triggered fruitful collaborations that have also led to the first book in this area. ISBN 978-3-540-25032-6.
- The follow-up Dagstuhl meeting (Seminar 07022, January 2007, Organizers: David Laidlaw and Joachim Weickert) was equally successful and the progress achieved is comprised in a second book published with Springer. ISBN 978-3-540-88377-7.
- The third Dagstuhl meeting (Seminar 09302, July 2009, Organizers: Bernhard Burgeth and David Laidlaw) paid special attention to engineering applications of tensors in fluid mechanics, material science, and elastography. It became apparent that these disciplines are facing many open problems in tensor field visualization and processing, and that the appropriate answers would greatly enhance the progress in these fields of engineering. The success of this meeting was documented by a third book authored by the participants. ISBN 978-3-642-27342-1.
- The fourth Dagstuhl meeting (Seminar 11501, December 2011, Organizers: Bernhard Burgeth, CF Westin and Anna Vilanova) witnessed a shift towards higher-order descriptors that went beyond tensors. Its numerous successful results are documented in a fourth Springer book. ISBN 978-3-642-54301-2.
- The fifth Dagstuhl meeting (Seminar 14082, February 2014, Organizers: Bernhard Burgeth, Ingrid Hotz, Anna Vilanova and CF Westin) has seen an increasing interest in statistical analysis and in the use of machine learning methods, as well as in the challenges posed by a novel generation of diffusion MR acquisition techniques. Its results were again collected in a Springer book. ISBN 978-3-319-15090-1.

6 Schedule

	Monday	Tuesday	Wednesday	Thursday	Friday
9:00	Introductions	Cheng Guan Koay Thomas Schultz	Bernhard Burgeth Hans Hagen	Evren Özarslan Andrada İnanç	Book
		Burak Acar	Jaya Sreevalsan Nair	Benoit Sherrer	
10:30	Coffee	Coffee	Coffee	Coffee	Coffee
11:00	Andrea Fuster Georgeta Elisabeta Marai	Luc Florack Tom Dela Haije	Aasa Feragen Xavier Tricoche	Anna Vilanova Bartroli Yue Zhang	Wrap-up
12:00	Lunch	Lunch	Lunch	Lunch	Lunch
13:30	Chunlei Liu Rodrigo Moreno Gözde Ünal	Ingrid Hotz Valentin Zobel Gerik Scheuermann	Social Event	Anders Eklund Alexandra Tobisch Maryam Afzali-Deligani	
15:00	Coffee	Coffee		Coffee	
15:30	Breakout 1 Elisabeta Marai	Breakout 2 Luc Florack		Breakout 4 Anna Vilanova Bartroli	
18:00	Dinner	Dinner		Dinner	
20:00		Breakout 3 Gerik Scheuermann			

Participants

- Burak Acar
Bogaziçi University –
Istanbul, TR
- Maryam Afzali-Deligani
Sharif University of Technology –
Tehran, IR
- Bernhard Burgeth
Universität des Saarlandes, DE
- Tom Dela Haije
TU Eindhoven, NL
- Anders Eklund
Linköping University, SE
- Aasa Feragen
University of Copenhagen, DK
- Luc Florack
TU Eindhoven, NL
- Andrea Fuster
TU Eindhoven, NL
- Hans Hagen
TU Kaiserslautern, DE
- Ingrid Hotz
Linköping University, SE
- Andrada Ianaş
University College London, GB
- Cheng Guan Koay
Walter Reed Medical Center –
Bethesda, US
- Chunlei Liu
University of California –
Berkeley, US
- Georgeta Elisabeta Marai
University of Illinois –
Chicago, US
- Rodrigo Moreno
KTH Royal Institute of
Technology – Stockholm, SE
- Jaya Sreevalsan Nair
IIIT – Karnataka, IN
- Jos B.T.M. Roerdink
University of Groningen, NL
- Evren Özarslan
Linköping University, SE
- Benoit Sherrer
Harvard Medical School –
Boston, US
- Geric Scheuermann
Universität Leipzig, DE
- Thomas Schultz
Universität Bonn, DE
- Alexandra Tobisch
DZNE – Bonn, DE &
Universität Bonn, DE
- Xavier Tricoche
Purdue University –
West Lafayette, US
- Gözde Ünal
Istanbul Technical University, TR
- Anna Vilanova Bartroli
TU Delft, NL
- Yue Zhang
Oregon State University, US
- Valentin Zobel
Universität Leipzig, DE



Foundations of Data Management

Edited by

Marcelo Arenas¹, Richard Hull², Wim Martens³, Tova Milo⁴, and Thomas Schwentick⁵

1 Pontificia Universidad Catolica de Chile, CL, marenas@ing.puc.cl

2 IBM TJ Watson Research Center – Yorktown Heights, US, hull@us.ibm.com

3 Universität Bayreuth, DE, wim.martens@uni-bayreuth.de

4 Tel Aviv University, IL, milo@cs.tau.ac.il

5 TU Dortmund, DE, thomas.schwentick@udo.edu

Abstract

In this Perspectives Workshop we have explored the degree to which principled foundations are crucial to the long-term success and effectiveness of the new generation of data management paradigms and applications, and investigated what forms of research need to be pursued to develop and advance these foundations.

The workshop brought together specialists from the existing database theory community, and from adjoining areas, particularly from various subdisciplines within the Big Data community, to understand the challenge areas that might be resolved through principled foundations and mathematical theory.

Perspectives Workshop April 10–15, 2016 – <http://www.dagstuhl.de/16151>

1998 ACM Subject Classification H.2 Database Management

Keywords and phrases Foundations of data management, Principles of databases

Digital Object Identifier 10.4230/DagRep.6.4.39

Edited in cooperation with Pablo Barceló

1 Executive Summary

Marcelo Arenas

Richard Hull

Wim Martens

Tova Milo

Thomas Schwentick

License  Creative Commons BY 3.0 Unported license
© Marcelo Arenas, Richard Hull, Wim Martens, Tova Milo, and Thomas Schwentick

The focus of Foundations of Data Management (traditionally termed Database Theory) is to provide the many facets of data management with solid and robust mathematical foundations. The field has a long and successful history and has already grown far beyond its traditional scope since the advent of the Web.

The recent push towards Big Data, including structured, unstructured and multi-media data, is transforming and expanding the field at an unusually rapid pace. However, for understanding numerous aspects of Big Data, a robust research exploration into the principled foundations is still lacking. This transformation will call upon the Database Theory community to substantially expand its body of tools, techniques, and focal questions and to much more fully embrace several other disciplines, most notably statistics and probability



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

Foundations of Data Management, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 39–56

Editors: Marcelo Arenas, Richard Hull, Wim Martens, Tova Milo, and Thomas Schwentick



DAGSTUHL
REPORTS

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

theory, natural language processing, data analytics, emerging hardware and software supports for computation, and data privacy and security.

Big Data is not the only force that is driving expansion and transformation for the Foundations of Data Management. With the increasing digitization of diverse industries, including “smarter cities”, education, healthcare, agriculture and others, many diverse kinds of data usage at large scales are becoming crucial. The push towards data-centric business processes, which are especially important for knowledge-worker driven processes, raise fundamentally new questions at the intersection of data and process. And increasing adoption of semantic web and other ontology-based approaches for managing and using meta-data push the boundaries of traditional Knowledge Representation.

The purpose of this Dagstuhl Perspectives Workshop was to explore the degree to which principled foundations are crucial to the long-term success and effectiveness of the new generation of data management paradigms and applications, and to understand what forms of research need to be pursued to develop and advance these foundations.

For this workshop we brought together specialists from the existing database theory community, and from adjoining areas, such as Machine Learning, Database Systems, Knowledge Representation, and Business Process Management, to understand the challenge areas that might be resolved through principled foundations and mathematical theory.

More specifically, during this workshop we worked on:

- Identifying areas, topics and research challenges for Foundations of Data Management in the forthcoming years, in particular, areas that have not been considered as Database Theory before but will be relevant in the future and of which we expect to have papers at PODS and ICDT, the main conferences in the field.
- Outlining the techniques that will be most fruitful as starting points for addressing the new foundational challenges in Data Management.
- Characterising the major challenge areas in Big Data where a principled, mathematically-based approach can provide important contributions.
- Finding research goals in neighbouring areas that may generate synergies with our own.

The workshop consisted of eight invited tutorials on selected topics: (1) Managing Data at Scale, (2) Uncertainty and Statistics in Foundations of Data Management, (3) Human in the Loop in Data Management, (4) Machine Learning and Data Management, (5) Data-Centric Business Processes and Workflows, (6) Ethical Issues in Data Management, (7) Knowledge Representation, Ontologies, and Semantic Web, and (8) Classical DB Questions on New Kind of Data. The abstracts of these talks can be found below in the document.

There were also seven working groups on theory-related topics, which identified the most relevant research challenges for Foundations of Data Management in the forthcoming years, outlined the mathematical techniques required to tackle such problems, and singled out specific topics for insertion in a curriculum for the area. The topics of these working groups were: (1) Imprecise Data, (2) Unstructured and Semi-structured Data, (3) Process and Data, (4) Data Management at Scale, (5) Data Management and Machine Learning, (6) Knowledge-Enriched Data Management, and (7) Theory and Society. There was also a working group on curriculum related issues, that collected and enriched the information provided by the working groups about the design of a curriculum on Foundations of Data Management. Each one of these groups worked for two consecutive hours in different days. Workshop participants had to participate in at least two working groups, although most of the people participated in four of them. Summaries of the discussions held in each one of these working groups can be found below in the document.

During the first day of the workshop, there were also five working groups that analysed several community-related aspects. In particular: (1) Attraction of women and young mem-

bers, (2) cross-fertilization with neighbouring areas, (3) relationship to industry, (4) impact of our research, and (5) the publishing process. The discussion within some of these working groups gave rise to the creation of specific tasks to be accomplished by the community in the following years. These tasks will be coordinated by the councils of PODS and ICDT, the two main conferences in the field.

This Dagstuhl Report will be accompanied by a Dagstuhl Manifesto, in which the outcome of the different working groups will be explained in more detail and several strategies for the development of our field will be proposed.

2 Table of Contents

Executive Summary

Marcelo Arenas, Richard Hull, Wim Martens, Tova Milo, and Thomas Schwentick 39

Overview of Talks

Issues in Ethical Data Management
Serge Abiteboul 43

(Non-)Classical DB Questions on New Kinds of Data
Marcelo Arenas and Pablo Barceló 43

Knowledge Representation, Ontologies, and Semantic Web
Georg Gottlob and Carsten Lutz 43

Machine Learning and Data Management
Eyke Hüllermeier 44

Uncertainty and Statistics in Foundations of Data Management
Benny Kimelfeld and Dan Suciu 44

Human-in-the-loop in Data Management
Tova Milo 44

Data Management at Scale
Dan Suciu and Ke Yi 45

Data-Centric Business Processes and Workflows
Victor Vianu and Jianwen Su 45

Working groups

Theory and Society
Serge Abiteboul 46

Knowledge-enriched Data Management
Diego Calvanese 47

Semistructured and Unstructured Data
Claire David 48

Data Management and Machine Learning
Eyke Hüllermeier 49

Imprecise data
Leonid Libkin 50

Curriculum for Foundations of Data Management
Frank Neven 51

Process and Data
Victor Vianu 52

Managing data at scale
Ke Yi 53

Participants 56

3 Overview of Talks

3.1 Issues in Ethical Data Management

Serge Abiteboul (ENS – Cachan, FR)

License © Creative Commons BY 3.0 Unported license
© Serge Abiteboul

In the past, database research was driven by data model and performance. In the future, personal/social data and ethics will. What should be done? Change how we deal with personal data? Change the web? We will consider various issues: Privacy, data analysis, data quality evaluation, data dissemination, and data memory.

3.2 (Non-)Classical DB Questions on New Kinds of Data

Marcelo Arenas (Pontificia Universidad Catolica de Chile, CL) and Pablo Barceló (University of Chile – Santiago de Chile, CL)

License © Creative Commons BY 3.0 Unported license
© Marcelo Arenas and Pablo Barceló

Different data models have been proposed in the last 20 years for representing semistructured or unstructured data. These include, e.g., graph databases, XML, RDF, JSON, and CSV. We explain that over such models we have (a) some classical DB questions being particularly important (e.g., schema design, query languages, and updates), (b) other classical DB questions gaining renewed interest (e.g., uncertainty, distribution, workloads, etc), and (c) some new questions appearing in full force (e.g., trust, access, variety, schema extraction).

Via examples we show that these problems are not only of theoretical interest, but, more importantly, that theory can have a strong positive impact in their practice. We start by presenting the work done in the standardization of a declarative query language for graph databases, which has concentrated the efforts of academics and practitioners in the last year. In this case theory has helped identifying a reasonable tradeoff between expressiveness and computational cost for the language. We then concentrate on the case of RDF and its query language SPARQL. We briefly review several practical problems that remain open regarding SPARQL evaluation in a distributed and open environment (the Web), and sketch the kind of theoretical understanding that is needed for their solution.

3.3 Knowledge Representation, Ontologies, and Semantic Web

Georg Gottlob (University of Oxford, GB) and Carsten Lutz (Universität Bremen, DE)

License © Creative Commons BY 3.0 Unported license
© Georg Gottlob and Carsten Lutz

The tutorial gave an overview of the recent developments in data access with ontologies. The first part focussed on the case where ontologies are formulated in a description logic (DL), surveying in particular the different families of DLs and their relations, as well as current research topics. The second part concentrated on the case where ontologies are formulated

in existential rule languages, which generalize Horn DLs and other relevant formalisms. It surveyed relevant language families and their relations, as well as recent results.

3.4 Machine Learning and Data Management

Eyke Hüllermeier (Universität Paderborn, DE)

License  Creative Commons BY 3.0 Unported license
© Eyke Hüllermeier

This tutorial-style presentation starts with a brief introduction to machine learning, specifically tailored for an audience with a background in database theory. The main part of the talk is then devoted to an overview of large-scale machine learning, that is, the application of machine learning methods to massive amounts of data, where scalability and efficient data management clearly become an issue. Finally, the talk also addresses the question of what machine learning can contribute to database management.

3.5 Uncertainty and Statistics in Foundations of Data Management

Benny Kimelfeld (Technion – Haifa, IL) and Dan Suciu (University of Washington – Seattle, US)

License  Creative Commons BY 3.0 Unported license
© Benny Kimelfeld and Dan Suciu

In this tutorial we outline the state of affairs on the research of managing probabilistic and statistical data, while focusing on the angle of database theory. We begin with an overview of past research, where we cover three main aspects: modeling probabilistic databases, complexity of probabilistic inference, and relevant applications. In particular, we describe established connections between query evaluation over tuple-independent databases, weighted model counting, and inference over Markov Logic Networks. We also discuss some key techniques such as lifted inference, tree decomposition, and knowledge compilation. Finally, we propose several directions for future research, where we make the case for a tighter incorporation of the sources of uncertainty (beyond the uncertainty itself) into the formal models and associated algorithms.

3.6 Human-in-the-loop in Data Management

Tova Milo (Tel Aviv University, IL)

License  Creative Commons BY 3.0 Unported license
© Tova Milo

Modern data analysis combines general knowledge stored in databases with individual knowledge obtained from the crowd, capturing people habits and preferences. To account for such mixed knowledge, along with user interaction and optimisation issues, data management platforms must employ a complex process of reasoning, automatic crowd-task generation and result analysis. This tutorial introduces the notion of crowd mining and describes a generic architecture for crowd mining applications. This architecture allows to examine and

compare the components of existing crowdsourcing systems and point out extensions required by crowd mining. It also highlights new research challenges and potential reuse of existing techniques/components. We exemplify this for the OASSIS project, a system developed in Tel Aviv University, and for other prominent crowdsourcing frameworks.

3.7 Data Management at Scale

Dan Suciu (University of Washington – Seattle, US) and Ke Yi (HKUST – Kowloon, HK)

License © Creative Commons BY 3.0 Unported license
© Dan Suciu and Ke Yi

The tutorial starts with a brief discussion of technology trends relevant to scalable data management research. Next, it presents four theoretical results on the complexity of multi-join query evaluation: the AGM bound and worst case sequential algorithms, the communication cost for parallel algorithm, the I/O cost in the external memory model, and recent sampling-based approximation algorithms. It ends with a brief discussion on potential topics to be included in a future course or book on Foundations of Data Management.

3.8 Data-Centric Business Processes and Workflows

Victor Vianu (University of California – San Diego, US) and Jianwen Su (University of California – Santa Barbara, US)

License © Creative Commons BY 3.0 Unported license
© Victor Vianu and Jianwen Su

A business process (BP) is an assembly of tasks to accomplish a business goal. A workflow is a software system managing business process executions. Traditional BP/workflow models do not include data, which makes it difficult to accurately specify processes in which data plays a central role, and raises many problems for their analysis, inter-operation, and evolution. This tutorial provides an introduction to the practical issues in data-centric BP/workflow management, and surveys existing theoretical results, addressing the following topics: (1) Formal analysis, verification, and synthesis of workflows, (2) workflow management system design, and (3) other workflow management issues including process mining and interoperation. We argue that the marriage of process and data is a fertile ground for new database theory questions, bringing into play a mix of techniques from finite-model theory, automata theory, computer-aided verification, and model checking. The new setting requires extending classical approaches to views, data exchange, data integration, expressiveness measures, and incomplete information.

4 Working groups

4.1 Theory and Society

Serge Abiteboul (ENS – Cachan, FR)

License  Creative Commons BY 3.0 Unported license
© Serge Abiteboul

(*Moderator:* Sege Abiteboul. *Participants:* Sudeepa Roy, Juan Reutter, Ronald Fagin, Jianwen Su, Thomas Schwentick, Stijn Vansummeren, Thomas Eiter, Iovka Boneva).

Society asks questions such as: (a) Who owns my medical data? (b) Is the information I receive of good quality? (c) Can Google influence the US president election? Theory asks by: (a) Access control (b) Tools for fair data analysis (c) Tools for defining and checking bias.

Regarding the first one, access control, it should consider questions such as “who can access sensitive data?” and “how is sensitive data being used?” In this context we think there is a need for developing:

- Models for access control over distributed data.
- Languages and user friendly tools for (i) specifying how the data can be used, (ii) understanding which data has been used for what, and (iii) control whether the data has been used in agreement with the access policy.

This might allow nice users to be guided in order to understand which queries are forbidden. It might also help detecting malicious users a posteriori.

Regarding the second one, fair data analysis, it might be of help for the data analyst (e.g., do I make good quality data analysis?) and the non-expert user (e.g., is this a relevant recommendation?). We think that in this context there is a need for developing:

- User friendly tools that check whether data analysis was fair.
- Helper tools for data analysts that check fairness criteria and guide in correcting the fairness errors.

Regarding the third one, bias and responsibility, this is of great importance not only for society but also for researchers (e.g., what can we do as researchers for designing less biased web services with high responsibility?). We think that in this context there is a need for developing:

- Political solutions (laws, control).
- Definitions of what it means having bias.
- Ranking algorithms that take into account the quality of pages.
- Technical tools for checking for the existence of bias, e.g., verification, testing, etc.

We also believe that several of these topics could be included in the curriculum of undergrad students, e.g., laws related to data, ethics (in general, for computer science), privacy issues and possible solutions, etc.

4.2 Knowledge-enriched Data Management

Diego Calvanese (Free University of Bozen-Bolzano, IT)

License  Creative Commons BY 3.0 Unported license
© Diego Calvanese

(*Moderator:* Diego Calvanese. *Participants:* Andreas Pieris, Carsten Lutz, Claire David, Filip Murlak, Georg Gottlob, Magdalena Ortiz, Marcelo Arenas, Meghyn Bienvenu, Paolo Guagliardo, Reinhard Pichler, Thomas Eiter, Jianwen Su, Ron Fagin, Sudeepa Roy).

The working group identified four important practical challenges:

1. Develop personalized and context-aware data access and management tools. Data in this case is highly heterogenous, multi-model and multi-modal. Here we deal with “small data” at the individual level, tuned to different view points.
2. Providing end users flexible and integrated access to large amounts of complex, distributed, heterogenous data (under different representations and different models). End users should be assumed to be domain experts, not data management experts.
3. Ensure interoperability at the level of systems exchanging data.
4. Bring knowledge to data analytics and data extraction.

It also identified seven relevant theoretical challenges related to the latter:

1. Development of reasoning-tuned DB systems, including new optimizations, new cost models, new/improved database engines optimized for reasoning, approximate answers, distributed evaluation, etc.
2. Choosing/designing the right languages for supporting these tasks. Here we need pragmatic choices motivated by user needs, but also supporting different types of knowledge and data (e.g., mixing CWA+OWA, temporal, spatial, etc.)
3. We need new measures of complexity for understanding easy/difficult cases, that explain better what works in practice. It would be interesting to explore alternative complexity measures, such as parameterized and average complexity, measuring complexity on the Web, smoothed analysis, etc.
4. Building user-friendly interfaces (beyond Protege). In particular, we need tools support geared towards end users (i.e., domain experts, lay people, and not just IT/knowledge engineers), but also support for query formulation, tools for exploration, etc.
5. Developing next-generation reasoning services. Here we need to explore notions related to explanation, abduction, hypothetical reasoning, defeasible reasoning, etc.
6. Reasoning with imperfect data, e.g., reasoning under contradictions and/or uncertainty, reasoning about quality of data, and support for improving data quality.
7. In depth study of temporal and dynamic aspects, such as managing changing data and knowledge, streaming data, reasoning about change, updating data in the presence of knowledge, etc.

4.3 Semistructured and Unstructured Data

Claire David (University Paris-Est – Marne-la-Vallée, FR)

License  Creative Commons BY 3.0 Unported license
© Claire David

(*Moderator:* Claire David. *Participants:* Iovka Boneva, Wim Martens, Juan Reutter Magdalena Ortiz, Domagoj Vrgoc, Filip Murlak, Frank Neven, Pablo Barcelo, Marcelo Arenas, Serge Abiteboul, Torsten Grust, Thomas Schwentick).

The huge amount of available data is perceived as a clear asset. However, exploiting this data meets several obstacles, as for instance the well known 3V: volume, velocity, variety of the data. One particular aspect of the variety of data is the co/existence of different formats for semi-structured and unstructured data, in addition to the widely used relational data format. A non exhaustive list is tree-structured data (e.g. XML, json), graph data (RDF and others), tabular data (e.g., CSV), temporal and spatial data, text, multimedia. We can expect that in the near future, new data formats will arise in order to cover particular needs.

The existence and coexistence of various formats is not new, but we believe that recent changes in the nature of available data raise a strong need for a new principled approach for dealing with different data models.

The database community has been working actively for many years now towards understanding each of these data formats formally by abstracting them, by developing good notions of schema, query languages, mappings and by studying the complexity of the related problems. These questions are still relevant and constitute important challenges.

Data heterogeneity is also an old problem in the database community, and until now has been tackled by data integration and data exchange solutions. A possible way for integrating data of multiple formats is to import all the data into a relational database, and use an RDMS afterwards for querying and managing the data; this solution is currently widely applied. The working group agreed that such a solution does have advantages, but does not fit all use cases; for instance, some data sources may contain dynamic data, and integration requires a lot of effort.

Additionally, the nature of available data has changed since the classical data integration and data exchange solutions have been proposed. The proliferation of many different data formats, the increase in the amount of available data, and the limited control users have on the data they are using, all challenge these solutions.

Therefore, we need a principled approach for dealing with different data models. For usability reasons, the new approach to be proposed must satisfy the following constraints:

- A.** It should be possible to keep the data in its original format, while still accessing data from different sources from a unique interface.
- B.** The approach should allow for adding new data models in the future.

The working group identified a number of problems that need to be solved in order to achieve this goal:

1. Understanding the data:
 - How to abstract and represent information about the structure of such data stored in various format? (Is there a good notion of multimodel schema? Can we use ontology?)
 - As open or third party data comes sometimes with no meta information available, we need tools for extracting such schema or at least partial structural information from data.
 - Develop entity resolution tools.

- Provide mapping between data in various formats.
 - Provide tools for extracting statistical properties of data, and tools for data analytic.
2. Accessing the data: There is a need for:
 - Specialized query languages for the different data models.
 - A general query language that can combine various specialized query languages.
 - Provide efficient algorithm for planning and evaluating a query using structural information of data.
 - Methods and tools for data summarization and data visualization.
 - Provide user friendly paradigms for both presenting the data, information about its structure and formulating queries.
 3. Orthogonal aspects:
 - Data in some applications can be highly distributed. How to handle distributed data processing, indexing, costs model?
 - Models for representing trust for data/information, privacy and data quality.
 - Do not forget the user who need usable tools to be able to understand and access the data.

4.4 Data Management and Machine Learning

Eyke Hüllermeier (Universität Paderborn, DE)

License  Creative Commons BY 3.0 Unported license
© Eyke Hüllermeier

(Moderator: Eyke Hüllermeier. Participants: Rick Hull, Floris Geerts, Benny Kimelfeld, Pablo Barcelo, Jens Teubner, Jan Van den Bussche, Stijn Vansummeren).

Machine Learning (ML) often focuses on large-scale learning, but several other directions exist that might be of interest to data management. These include: (a) Logic: Learning of formulas, concepts, ontologies, and so on (work in database theory and description logics), (b) Automata: Learning automata (which might have applications for learning XML schemas, for instance), (c) Workflows: Process mining, process discovery, and others, and (d) Query languages: Extensions of languages used in DBMS by ML functionalities (e.g., query-by-example), ILP, declarative data mining, etc.

We believe that there are several ML problems in which a data management perspective could be of help. These include:

1. Applying data management techniques in the optimization of ML algorithms, e.g., implementation of operators for data access, sequential vs parallel sampling, etc.
2. A finer complexity analysis of ML algorithms, in particular, their complexity in terms of different parameters such as the number of features, labels, label density, etc.
3. Studying I/O complexity of ML algorithms in external memory.
4. Building models for distributed ML based on frameworks such as MapReduce.

We also believe that data management could benefit from ML. For example:

1. Building query languages for DBMS that incorporate ML functionalities. These can be of importance for optimization, cleaning, analysis, etc.
2. Use of deep networks for semantic indexing/hashing.
3. Learning hash functions.
4. Predictive modeling, performance prediction.
5. Learning approximate dependencies, concepts, ontologies, etc.

4.5 Imprecise data

Leonid Libkin (*University of Edinburgh, GB*)

License  Creative Commons BY 3.0 Unported license
© Leonid Libkin

(*Moderator: Leonid Libkin. Participants: Meghyn Beinvenu, Angela Bonifati, Diego Calvanese, Paolo Guagliardo, Benny Kimelfeld, Phokion Kolaitis, Maurizio Lenzerini, Carsten Lutz, Tova Milo, Dan Olteanu, Sudeepa Roy, Dan Suciu, Jan Van den Bussche, Victor Vianu.*)

Incomplete, uncertain, and inconsistent information is ubiquitous in data management applications. This was recognized already in the 1970s, and since then the significance of the issues related to incompleteness/uncertainty has been steadily growing: it is a fact of life that data we need to handle on an everyday basis is rarely complete.

However, while the data management field developed techniques specifically for handling incomplete data, their current state leaves much to be desired. Even evaluating SQL queries over incomplete databases – a problem one would expect to be solved after 40+ years of relational technology – one gets results that make people say “*you can never trust the answers you get from [an incomplete] database*” (Date). And while even such basic problems remain unsolved, we now constantly deal with more varied types of incomplete and inconsistent data. There is thus an urgent need to address these problems from a theoretical point of view, keeping in mind that theoretical solutions must be usable in practice (indeed, this is the field where perhaps too much theoretical work focused on proving various impossibility results – intractability, undecidability – rather than addressing what can actually be done efficiently).

The main challenges are split into three groups.

Modeling uncertainty: This includes several themes, for example: (a) What are types of uncertainty we need to model/understand? (b) How do we store/represent uncertain information? What standard RDBMSs give us, for example, is very limited. (c) When can we say that some data is true? This issue is particularly relevant in crowdsourcing applications: having data that looks complete does not yet mean it is true. (d) How do we rank uncertain query answers? There is a tendency to divide everything into certain and non-certain answers, but this is often too coarse.

Reasoning with uncertainty There is much work on this subject but we want to address points close to data management: (a) How do we do inferences with incomplete data? (b) How do we integrate different types of uncertainty? (c) How do we learn queries on uncertain data? (d) What do query answers actually tell us if we run queries on data that is uncertain (that is, how results can be generalized from a concrete incomplete data set)?

Making it practical This is the most challenging direction. For far too long, theoretical literature identified small classes where queries behave well over incomplete data (often various classes of conjunctive queries) and then concentrated on proving intractability results outside those classes. We need to move on, and look at questions like those below: (a) How do we find good quality query answers even when we have theoretical intractability? For instance, how do we find answers with some correctness guarantees, and do so efficiently? (b) How do we make commercial RDBMS technology work well and efficiently in the presence of incomplete data? Even query optimization in this case is hardly a solved problem. (c) How do we make handling inconsistency (in particular,

consistent query answering) work in practice? How do we use it in data cleaning? (d) How do we achieve practically feasible query evaluation on probabilistic data?

Regarding the question of what problems do we need to solve soon. We believe that while all the above are important research topics that need to be addressed, it appears that there are several that can be viewed as a priority, not least because there is an immediate connection between theory and practice. This gives us a chance to make a good theory to solve practically relevant problems, and to make this theoretical work visible.

1. Can we fix the standard relational technology so that at least textbook writers would stop saying (justifiably!) that “*you’re getting wrong answers to some of your queries*”. We need to understand what it means to be right/wrong, and how to adjust the technology to ensure that wrong answers do not appear.
2. What should we use as benchmarks when working with incomplete/uncertain data? Quite amazingly, this has not been addressed; in fact standard benchmarks tend to just ignore incomplete data.
3. How do we devise approximation algorithms for classes of queries known to be intractable? The field is too dependent on producing results for conjunctive queries and leaving the rest for proving high-complexity results, but non-conjunctive queries do exist and need to be handled.
4. Is there any hope of making consistent query answering practical, and relevant (perhaps in data cleaning). Again too much emphasis was on proving dichotomies, even within conjunctive queries, and not making it work.

Regarding neighboring communities, we believe that our closest neighbors are database systems people (in terms of conferences, SIGMOD, VLDB). There are also interesting overlaps with things happening in uncertainty in AI (conferences such as UAI or SUM) and general AI conferences (KR, AAAI, IJCAI).

4.6 Curriculum for Foundations of Data Management

Frank Neven (Hasselt University – Diepenbeek, BE)

License © Creative Commons BY 3.0 Unported license
© Frank Neven

(*Moderator:* Frank Neven. *Participants:* Phokion Kolaitis, Thomas Schwentick, Leonid Libkin, Torsten Grust, Wim Martens, Pablo Barcelo, Reinhard Pichler, Marcelo Arenas, Stijn Vansummeren, Juan Reutter).

The working group on a curriculum for Foundations of Data Management (FDM) focused on the following questions. While the main goal of the group was to identify a curriculum for an FDM course on graduate level, the group also explored the relationship with other courses. A driving principle followed by the group was that we should not only prepare the next generation of PhD students in database theory (educate), but also expose the beauty and variety of our field to attract future PhD students (attraction) The group considered the following questions:

1. What are relevant concepts from FDM that could be featured as examples in non-database courses? The group gave the following examples of concepts: (a) trees can be related to XML, (b) graphs to the semantics Web, (c) regular expressions to querying of graphs

- and to content models in XML Schema, (d) mathematical structures in a logic course as examples of databases, and (e) laws related to data in an ethics in CS course.
2. What are the relevant concepts from FDM that can be featured in a first DB course (undergraduate level)? The following examples were given:
 - Data models: Relational, semi-structured, graphs.
 - Relational calculus as an abstraction of a query language.
 - Recursion and Datalog (as there is not much syntax involved here, you can quickly explain it and even use a system like LogicBlox to bring the concept in practice).
 - Well-designedness, constraints, normalization / BCNF.
 - The concept that data often doesn't fit into main memory (memory hierarchy / external memory).
 - Concurrency (an example here is a proof of the fact that 2-phase locking works).
 - The concept of incomplete information (e.g., the pitfalls of NULLS in SQL).
 3. What are the relevant topics for a graduate course in FDM? As a whole Dagstuhl seminar could be devoted to just this question, the following should just be seen as a start of discussion and would need further input from the community. Possible core topics that at the moment are not treated in depth in the AHV book:
 - Refined ways for studying complexity of queries.
 - Acyclic queries / Structural decomposition / tree decomposition (parametrized complexity).
 - AGM bound; join algorithms; leapfrog.
 - Data exchange and integration / relationship to KR.
 - Data provenance.
 - Graph data, RDF / Tree data, XML, JSON (tree automata).
 - Approximate query answering, top-k.
 - Stream-based processing.
 - External memory.
 - Incomplete information / Probabilistic databases.
 - Possible emerging topics: Map-reduce, parallel computation models.
 4. How should we proceed to write a book on FDM? It is safe to say that there was no general agreement on how to proceed. One possibility is to organize the book TATA-style where an editorial board is established that decides the topics, invites writers and ensures global consistency between the material. The effort could start from the AHV book by adapting and adding chapters.

4.7 Process and Data

Victor Vianu (University of California – San Diego, US)

License  Creative Commons BY 3.0 Unported license
© Victor Vianu

(Moderator: Victor Vianu. Participants: Thomas Schwentick, Thomas Eiter, Daniel Deutch, Magdalena Ortiz, Diego Calvanese, Jianwen Su, Richard Hull, Wim Martens, Serge Abiteboul).

Traditionally, workflow models have been process centric. Such a workflow specifies the valid sequencing of tasks by some finite-state transition diagram. This is inadequate in applications where data plays a central role, and has led to new workflow models in

which data is treated as a first-class citizen. The marriage of process and data represents a qualitative leap analogous to that from propositional logic to first-order logic.

Participants to the working group identified practical and technical challenges raised by data-centric workflows. The following were identified as important practical challenges:

1. Evolution and migration of business processes.
2. Automating manual processes (including workflow-on-the-fly).
3. Business Process compliance, correctness, non-functional properties, etc.
4. Business Process interaction and interoperation.
5. Business Process discovery and understanding (including analytics).
6. Workflow and business process usability (including human collaboration).

Tackling the above challenges raises a set of technical problems enumerated below (next to each are listed the relevant practical challenges):

- A. Verification and Static Analysis (relevant to 1,2,3,4).
 - Approximate verification via abstraction.
 - Incremental verification under workflow or property modifications.
 - Enabling/verifying “plug and play” of process building blocks.
 - Checking compliance (financial, government, security).
- B. Tools for Design and Synthesis (relevant to 1,2,4,6).
 - Operators for creating/modifying process schemas.
 - Full or partial synthesis from requirements and/or goals.
 - Inferring process dynamically from “digital exhaust”.
- C. Models and semantics for views, interaction, and interoperation (relevant to 1,2,3,4,5,6).
 - Providing customized views of business process schemas and instances.
 - Consistent usage of shared data and resources by multiple processes.
 - Views and Composition (e.g., orchestration, choreography, etc).
 - Inferring/guaranteeing properties of process compositions.
- D. Analytics with Process and Data (relevant to 1,2,3,4,5,6).
 - “Business Intelligence” – analytics over executions of business processes.
 - Process mining, process discovery.
 - Automatic identification of best practices.
 - Providing guidance and explanation at run time.

The participants also agreed that the lack of real data presents an obstacle to research in the area and that developing closer connections to industry would be beneficial. It is also desirable to explore further areas of applicability of this research, such as social networks or the blockchain approach used in financial transactions.

4.8 Managing data at scale

Ke Yi (HKUST – Kowloon, HK)

License © Creative Commons BY 3.0 Unported license
© Ke Yi

(Moderator: Ke Yi. Participants: Floris Geerts, Reinhard Pichler, Jan van den Bussche, Frank Neven, Filip Murlak).

In this working group we started by identifying different computational models that have been proposed for managing data at scale and some others that might be important to consider:

- **PRAM:** It has had some beautiful theory, but made little practical impact, reasons including 1) too far from reality, 2) hard to program. Nevertheless, there are many algorithmic ideas that are still very useful in other parallel models.
- **External memory:** Considered as a success, used as the standard model for DBMS algorithms. A related model is the cache-oblivious model which tries to model multiple levels of memory hierarchy. It is an elegant model, with some implementation work. But did not gain enough traction due to complication of the algorithms.
- **Streaming model:** Another successful model, with both nice theory and practical systems. More importantly, it has nurtured data sketching (data summarization in general), which is not only useful for data streams, but as a general approach in making big data small.
- **BSP, MapReduce, MPC, Pregel:** The original BSP model did not get enough attention due to complication: it had too many parameters and tried to model various costs. A computational model has to be both simple and realistic in order to be useful. The MPC model simplified the model and seems to hit the right simplicity-reality tradeoff. There is much active research, and more is expected to come.
- **Asynchronous models:** BSP, MPC etc are all synchronous models, but asynchronous models have not been studied fully in the database theory community (better studied in the distributed computing community).
- **Computational models for new hardware:** There are many practical researches in using GPU for speeding up database operations, but little theoretical work due to various architectures. CUDA might be a good model for more theoretical work. There are also many new chip designs that may be worth of looking at.
- **Relationship between the models:** Are there any general reductions between these models? Essentially, all these models are about locality and parallelism. If there is no general reduction, maybe there can be one for a class of algorithms or a class of problems (e.g., CQ, relational algebra).

We also identified several problems, where scalability is a major concern, that it would be relevant for the DB theory community to tackle:

- **Standard DB problems:** Of course, we should continue to work on standard DB problems (relational algebra, in particular joins, semi-structured data, XML, etc). Meanwhile, we should pay attention to the theory-practice gap.
- **Graph problems:** There is a lot of work on page rank, shortest path, mining, etc. Still many more needs to be solved at scale.
- **Machine learning at scale:** Machine learning is becoming more data-intensive. Linear algebra is a basic toolbox that needs to be scalable. New paradigms that are related to this community include distributed machine learning and learning on streams.
- **Transactions:** More theoretical work needs to address transactions. Issues like strong consistency and eventual consistency. Interaction with the distributed computing community might be desirable.
- **Scientific computing:** In particular large-scale simulation, which is becoming a dominating approach for scientific research.

We also identified some key techniques that these problems might require for their solution:

- **Beyond-worst-case analysis:** Worst-case analysis is still the first step in understanding the complexity of a problem, but may not be really relevant to an algorithm's practical performance. Many other angles should be explored, including, e.g., instance optimality, parameterized complexity, and smoothed complexity.

- **Algorithms specifically tailored for big data:** Just saying that a problem is in PTIME may not be enough for big data. We'd like to have algorithms that run in linear-time, near-linear time, or even sub-linear time. Parallel complexity is another important measure.
- **Approximation algorithms:** As data gets big, exact answers are not often required, and approximation algorithms can offer great efficiency speedup in this case. Sampling, MCMC, stratified sampling (BlinkDB), sampling over joins (by random walks) and many statistical techniques can be applied. Sublinear-time algorithms have been extensively studied in the algorithms community, but not fully exploited yet in DB.
- **Convex optimization:** In particular primal-dual methods, which have been recently applied for analyzing the maximum join size (AGM). It would be interesting to see if this can leave to more applications for DB questions.
- **Information theory:** We think it can be useful in proving lower bounds. It has been extensively used in the TCS community (communication complexity, etc), with a few examples in DB (communication lower bound for MPC model). We expect more applications to come.

Participants

- Serge Abiteboul
ENS – Cachan, FR
- Marcelo Arenas
Pontificia Universidad Catolica
de Chile, CL
- Pablo Barcelo
University of Chile – Santiago de
Chile, CL
- Meghyn Bienvenu
University of Montpellier, FR
- Iovka Boneva
Université de Lille I, FR
- Angela Bonifati
University Claude Bernard –
Lyon, FR
- Diego Calvanese
Free Univ. of Bozen-Bolzano, IT
- Claire David
University Paris-Est –
Marne-la-Vallée, FR
- Daniel Deutch
Tel Aviv University, IL
- Thomas Eiter
TU Wien, AT
- Ronald Fagin
IBM Almaden Center –
San Jose, US
- Floris Geerts
University of Antwerp, BE
- Georg Gottlob
University of Oxford, GB
- Torsten Grust
Universität Tübingen, DE
- Paolo Guagliardo
University of Edinburgh, GB
- Eyke Hüllermeier
Universität Paderborn, DE
- Richard Hull
IBM TJ Watson Res. Center –
Yorktown Heights, US
- Benny Kimelfeld
Technion – Haifa, IL
- Phokion G. Kolaitis
University of California –
Santa Cruz, US
- Maurizio Lenzerini
Sapienza University of Rome, IT
- Leonid Libkin
University of Edinburgh, GB
- Carsten Lutz
Universität Bremen, DE
- Wim Martens
Universität Bayreuth, DE
- Tova Milo
Tel Aviv University, IL
- Filip Murlak
University of Warsaw, PL
- Frank Neven
Hasselt Univ. – Diepenbeek, BE
- Dan Olteanu
University of Oxford, GB
- Magdalena Ortiz
TU Wien, AT
- Reinhard Pichler
TU Wien, AT
- Andreas Pieris
TU Wien, AT
- Juan L. Reutter
Pontificia Universidad Catolica
de Chile, CL
- Sudeepa Roy
Duke University – Durham, US
- Thomas Schwentick
TU Dortmund, DE
- Jianwen Su
University of California –
Santa Barbara, US
- Dan Suciu
University of Washington –
Seattle, US
- Jens Teubner
TU Dortmund, DE
- Jan Van den Bussche
Hasselt University, BE
- Stijn Vansummeren
Université Libre de Bruxelles, BE
- Victor Vianu
University of California –
San Diego, US
- Domagoj Vrgoc
Pontificia Universidad Catolica
de Chile, CL
- Ke Yi
HKUST – Kowloon, HK



Tensor Computing for Internet of Things

Edited by

Evrin Acar¹, Animashree Anandkumar², Lenore Mullin³,
Sebnem Rusitschka⁴, and Volker Tresp⁵

- 1 University of Copenhagen, DK, evrim@life.ku.dk
- 2 University of California – Irvine, US, a.anandkumar@uci.edu
- 3 University of Albany – SUNY, US, lmullin@albany.edu
- 4 Siemens AG – München, DE, sebnem.rusitschka@siemens.com
- 5 Siemens AG – München, DE, volker.tresp@siemens.com

Abstract

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 16152 “Tensor Computing for Internet of Things”. In an interactive three-day workshop industrial and academic researchers exchanged their multidisciplinary perspectives through impulse talks, panel discussions, and break-out sessions. Internet of Things (IoT) or Cyber-physical systems (CPS) bring out interesting new challenges to tensor computing, such as the need for real-time analytics and control in interconnected dynamic networks, e.g. electricity, transportation, manufacturing. On the other hand, IoT/CPS have characteristics that make tensor methods applicable to extract information very efficiently. During our discussions we identified an action plan to have a structured approach that will enable the multidisciplinary community of domain and control experts, data scientists, and distributed, embedded software developers to share knowledge and best practices, compare and exchange tensor models depending on data types and applications in distinct IoT/CPS scenarios.

Perspectives Workshop April 10–13, 2016 – <http://www.dagstuhl.de/16152>

1998 ACM Subject Classification D.4.7 Organization and Design, D.4.5 Reliability, I.2.6 Learning, I.1 Symbolic and Algebraic Manipulation, I.2.11 Distributed Artificial Intelligence

Keywords and phrases Tensor Methods, Multi-way Data Analysis, Multi-linear Algebra, Tensor Software, Distributed & Parallel Computing, Big Data Computing & Analytics, Cyber-physical Systems, Intelligent Autonomous Systems, Applications in Smart Grid, Mobility, Smart City

Digital Object Identifier 10.4230/DagRep.6.4.57

1 Executive Summary

Evrin Acar

Animashree Anandkumar

Lenore Mullin

Volker Tresp

Sebnem Rusitschka

License © Creative Commons BY 3.0 Unported license

© Evrin Acar, Animashree Anandkumar, Lenore Mullin, Volker Tresp, and Sebnem Rusitschka

In April 2016, Dagstuhl hosted a Perspectives Workshop on Tensor Computing for the Internet of Things. The prior year, industrial researchers had formulated the challenges of gaining insights from multi-dimensional sensory data coming from large-scale connected



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

Tensor Computing for Internet of Things, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 57–79

Editors: Evrin Acar, Animashree Anandkumar, Lenore Mullin, Sebnem Rusitschka, and Volker Tresp



Dagstuhl Reports

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

energy, transportation networks or manufacturing systems. The sheer amount of streaming multi-aspect data was prompting us to look for the most suitable techniques from the machine learning community: multi-way data analysis. Hence, we organized a three-day interactive workshop with two separate questions bringing two formerly distinct communities together: (i) How can we assure performance and reliability given the increasing complexity and data of an always-on connected world? (ii) Can we exploit the power of tensor algebra to solve high-dimensional large-scale machine learning problems that such a world poses?

The workshop focused on the Internet of Things (IoT), i.e. devices, which have the capability to sense, communicate, and more so, control their environments. These devices are increasingly becoming a part of complex, dynamic, and distributed systems of electricity or mobility networks, hence our daily lives. Various sensors enable these devices to capture multiple aspects of their surroundings in real-time. For example, phasor measurement units capture transient dynamics and evolving disturbances in the power system in high-resolution, in a synchronized manner, and in real-time. Another example is traffic networks, where a car today can deliver about 250 GB of data per hour from connected electronics such as weather sensors within the car, parking cameras and radars. Experts estimate that the IoT will consist of almost 50 billion objects by 2020 [2], which will trigger the Era of Exascale computing necessitating the management of heat and energy of computing in concert with more and more complex processor/network/memory hierarchies of sensors and embedded computers in distributed systems. Crucial for the extraction of relevant information is the format in which the raw data from such systems is represented. Crucial for viable efficiency of information extraction in IoT is which operations are used guaranteeing various attributes of resource use and management. Tensors can be viewed as data structures or as multilinear operators.

The goal of the workshop was to explore tensor representations and computing as the basis for machine learning solutions for the IoT. Tensors are algebraic objects which describe linear and multilinear relationships, and can be represented as multidimensional arrays. They often provide a natural and compact representation for multidimensional data. In the recent years, tensor and machine learning communities – mainly active in the data-rich domains such as neuroscience, social network analysis, chemometrics, knowledge graphs etc. – have provided a solid research infrastructure, reaching from the efficient routines for tensor calculus to methods of multi-way data analysis, i.e., tensor decompositions, to methods for consistent and efficient estimation of parameters of the probabilistic models.

Some tensor-based models have the intriguing characteristic that if there is a good match between the model and the underlying structure in the data, the models are much better interpretable than alternative techniques. Their interpretability is an essential feature for the machine learning techniques to gain acceptance in the rather engineering heavy fields of automation and control of cyber-physical systems. Many of these systems show intrinsically multilinear behavior, which is appropriately modeled by tensor methods and tools for controller design can use these models. The calibration of sensors delivering data and the higher resolution of measured data will have an additional impact on the interpretability of models.

Various presentations on tensor methods by established researchers from different application domains underscored that tensor methods are reaching a maturity tipping point. However, knowledge of usage characteristics of tensor models is scattered. Discussions of the currently independent perspectives on the usage of tensor methods showed convergence potential which we will detail in the Dagstuhl Manifesto. During our discussions based on the presentations of the IoT industrial researchers, it quickly became clear that we would

need benchmark challenges for cyber-physical systems and benchmark data in order to be able to replicate the successes in machine learning for neuroscience, image processing or chemometrics, for example.

The tensor computing community will equally benefit from the new types of data, requirements and characteristics of IoT, which can lead to techniques that increase success rates of previous applications, as was the case with the challenges of social network data analysis leading to better tensor models/algorithms that can analyze data sets with missing entries, now used in many other fields in addition to social network analysis. Additionally, as opposed to standardized machine learning techniques, tensor computing currently lacks a common language and the homogeneity to flexibly exchange models. Hence, a hub platform bringing data and domain knowledge of cyber-physical systems together with a variety of practitioners of tensor computing would enhance increasing coherence of terms, best practices in data acquisition and structuring methods as well as model benchmarking, cataloging, and exchange of methods.

Furthermore, industrial researchers from IoT, automation and control domains highlighted their view that tensor computing methods are currently still inaccessible to the majority of the industrial practitioners even though there has been a considerable progress in developing tools for tensor computing. Matlab extensions to enable the use of tensor analysis are quite mature [1] [3] [4]. Matlab is widely used by control and automation practitioners. Python ecosystem for machine learning practitioners is very quickly adopting extensions for enabling tensor operations [5] [6]. However, both are mainly for prototyping and ultimately do not fulfill the need for a unified framework for industrial grade development and deployment of models in highly distributed cyber-physical systems. Interestingly, just five months prior to our workshop, Tensorflow [7], a numerical computation library aiming at capturing structures in multidimensional data as well as supporting both prototyping and production level algorithms was open sourced. Tensorflow can run on server clusters as well as embedded systems such as smart phones [8]. Another framework, unifying both batch and streaming data analysis, is Apache Spark [9]. Spark provides seamless scalability of software code to run on multiple machines. Recently there have been deployments of tensor methods on the Spark platform.

As a multidisciplinary community we believe that we will be able to formulate requirements and provide support in developing improvements for unifying frameworks. The required skill set is quite rare: we are in need of software developers that can create reliable high-performant code for both server-side distributed training on massive amounts of data and deployment of trained models in embedded distributed system. Heterogeneous processor architectures are predominant in cyber-physical systems. Either these software developers should be data scientists proficient in tensor computing and very good at communicating with domain experts or we need tooling such that data scientists and domain experts can collaboratively model data for cyber-physical systems. We will detail these discussions in the Manifesto: We believe that it is feasible to create such tooling that automates the generation of reliable, secure code, which accounts for the adaptive logic of devices interacting with their dynamic physical environment – but also through which there is a direct feedback between data scientist, domain or control expert, and the adaptive control device.

The Manifesto, which will be published on www.dagstuhl.de/16152/ will include a roadmap of how we as a newly formed multidisciplinary community want to start with a knowledge hub on tensors, and iterate through data grand challenges from IoT pilots, results dissemination, into what may one day become collaborative modeling hub for learning cyber-physical systems.

References

- 1 Tamara G. Kolda Brett W. Bader et al. Matlab tensor toolbox version 2.6. <http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html>
- 2 Statista. Internet of things (iot): number of connected devices worldwide from 2012 to 2020 (in billions). <http://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/>
- 3 Ivan Oseledets. Tt (tensor train) toolbox version 2.2.2. <https://github.com/oseledets/TT-Toolbox>
- 4 Laurent Sorber Marc Van Barel Nico Vervliet, Otto Debals and Lieven De Lathauwer. Tensor lab user guide, release 3.0. <http://www.tensorlab.net/userguide3.pdf>
- 5 Ivan Oseledets. Python implementation of the tt-toolbox. <https://github.com/oseledets/ttpython>
- 6 Maximilian Nickel. scikit-tensor: Python library for multilinear algebra and tensor factorizations. <https://github.com/mnick/scikit-tensor>
- 7 Google. Tensorflow: An open source software library for numerical computation using data flow graphs. <https://www.tensorflow.org/>
- 8 Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zhang. Tensorflow: A system for large-scale machine learning. *CoRR*, abs/1605.08695, 2016
- 9 Apache Spark. Apache spark: A fast and general engine for large-scale data processing. <http://spark.apache.org/>

2 Table of Contents

Executive Summary

<i>Evrin Acar, Animashree Anandkumar, Lenore Mullin, Volker Tresp, and Sebnem Rusitschka</i>	57
--	----

Overview of Talks

IoT and Applications such as Predictive Maintenance <i>Christine Preisach (SAP SE – Walldorf, DE)</i>	63
Supervisory Control & Fault Diagnosis – Tensors in a Networking World <i>Gerwald Lichtenberg (HAW – Hamburg, DE)</i>	63
Heterogeneous Computing Infrastructures in Future Energy and Transportation Networks <i>Sebnem Rusitschka (Siemens AG – München, DE)</i>	64
Tensor analysis for handling huge amounts of -omics data <i>Rasmus Bro (University of Copenhagen, DK)</i>	64
Tensor Applications in Neuroscience <i>Morten Mørup (Technical University of Denmark – Lyngby, DK)</i>	64
Data Fusion using Coupled Matrix and Tensor Factorizations <i>Evrin Acar (University of Copenhagen, DK)</i>	65
Tensors for Representational Learning <i>Volker Tresp (Siemens AG – München, DE)</i>	66
Advances in the numerical computation of (coupled) tensor decompositions <i>Lieven De Lathauwer (KU Leuven, BE)</i>	66
Tensors and IoT Computational Issues <i>Lenore Mullin (University of Albany – SUNY, US)</i>	66
Guaranteed Learning Using Tensor Methods <i>Animashree Anandkumar (University of California – Irvine, US)</i>	67
Tensor techniques for the visualization of multidimensional data <i>Renato Pajarola (Universität Zürich, CH)</i>	68

Panel discussions

Distributed PARAFAC for Large Industrial Internet-scale Dense Tensors with Skewed Modes <i>Kareem Aggour (General Electric – Niskayuna, US)</i>	68
Sensors in Power Networks – Representing Network Dynamics over Space and Time <i>Denis Krompaß (Siemens AG – München, DE)</i>	69
Loose Semantic Coupling in IoT and the Role of Tensors <i>Edward Curry for Souleiman Hasan (National University of Ireland – Galway, IE)</i>	69
Key Aspects in Tensor Computing in the Context of IoT <i>Benoit Meister (Reservoir Labs, Inc. – New York, US)</i>	70
Modeling of Complex, Man-made Systems <i>Bülent Yener (Rensselaer Polytechnic Institute – Troy, US)</i>	70

Future Research Directions on Tensor Computing for IoT <i>Vagelis Papalexakis (Carnegie Mellon University, US)</i>	71
Working groups	
Applications of Tensor Decomposition in IoT Scenarios	71
Frameworks for Tensor Computing in IoT	73
Open Problems	77
Outlook	77
Participants	79

3 Overview of Talks

Impulse talks were grouped into two sessions: “IoT Applications and Computing Infrastructures” by industrial IoT/CPS researchers, who gave an overview and some deep dives into the IoT applications, multidimensional IoT data, what information we want to extract from it for what purpose, data sources and peculiarities, IoT/CPS computing infrastructures and peculiarities. “Challenges of Tensor Computing for IoT” consisted of impulse talks by academic researchers with vast experience in applying tensor decompositions, available tools, and emerging computing paradigms.

Impulse Talks I: IoT Applications and Computing Infrastructures

3.1 IoT and Applications such as Predictive Maintenance

Christine Preisach (SAP SE – Walldorf, DE)

License © Creative Commons BY 3.0 Unported license
© Christine Preisach (SAP SE – Walldorf, DE)

In this presentation Predictive Maintenance, one major application in IoT, is described using an example use case and applied data mining methods. Moreover challenges encountered in the data used for Predictive Maintenance are stated.

3.2 Supervisory Control & Fault Diagnosis – Tensors in a Networking World

Gerwald Lichtenberg (HAW – Hamburg, DE)

License © Creative Commons BY 3.0 Unported license
© Gerwald Lichtenberg (HAW – Hamburg, DE)

In this impulse talk “Internet of Things” is approached from a systems and control theory perspective. Interesting challenges of IoT include highly connected control structures, in which now non-experts not only have access to sensors but also to actors in the system. Challenging scientific questions are briefly addressed such as performance and robustness in heterogeneous networks, finding optimal solutions for fault diagnosis and supervision, and deriving useful models with structural properties from continuous and discrete data [1]. Tensors could help in investigating these questions, by utilizing the inherent multilinearity of continuous and discrete dynamics, the model structure of networks and systems, among other characteristics [2].

References

- 1 Thorsten Müller, Kai Kruppa, Gerwald Lichtenberg, and Nicolas Réhault. Fault detection with qualitative models reduced by tensor decomposition methods. *IFAC-PapersOnLine*, 48(21):416–421, 2015
- 2 Georg Pangalos, Annika Eichler, and Gerwald Lichtenberg. Hybrid multilinear modeling and applications. In *Simulation and Modeling Methodologies, Technologies and Applications*, pages 71–85. Springer, 2015.

3.3 Heterogeneous Computing Infrastructures in Future Energy and Transportation Networks

Sebnem Rusitschka (Siemens AG – München, DE)

License  Creative Commons BY 3.0 Unported license
© Sebnem Rusitschka (Siemens AG – München, DE)

Our world becomes increasingly computerized, “we put little computers in our ears to hear better, we put our bodies into computers that drive,” traditional sectors like energy, transportation, manufacturing increasingly rely on computerized automation [1].

The computing resources we’ll find in these systems are heterogeneous, with highly differing capabilities and restrictions, some instructions will be time-critical, safety-critical. So the applications/algorithms, which will learn from the data, running in these heterogeneous computing environments will need end-to-end support to verifiably execute in the most optimal and safe way.

Can we facilitate this? How long will it take?

References

- 1 Sebnem Rusitschka and Edward Curry. Big data in the energy and transport sectors. In *New Horizons for a Data-Driven Economy*, pages 225–244. Springer, 2016

Impulse Talks II: Tensor Decompositions

3.4 Tensor analysis for handling huge amounts of -omics data

Rasmus Bro (University of Copenhagen, DK)

License  Creative Commons BY 3.0 Unported license
© Rasmus Bro (University of Copenhagen, DK)

In this talk I will show how tensor methods are crucial in understanding the bigger and bigger datasets obtained on health, environmental and food research. With tensors methods we are able to convert huge and difficult chemical measurements into smaller data sets that are easier to handle, present the information in a more ‘correct’ manner and helps eliminating spurious findings which often arise from mining huge unfiltered data [1].

References

- 1 Age Smilde, Rasmus Bro, and Paul Geladi. *Multi-way Analysis: Applications in the Chemical Sciences*, West Sussex: Wiley, 2004.

3.5 Tensor Applications in Neuroscience

Morten Mørup (Technical University of Denmark – Lyngby, DK)

License  Creative Commons BY 3.0 Unported license
© Morten Mørup (Technical University of Denmark – Lyngby, DK)

In this talk a range of applications of tensor decomposition in neuroscience will be outlined in particular for the modeling of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) data. These spatio-temporal data of brain activity naturally form tensors when measured across multiple subjects, trials or conditions. Modeling approaches

based on the CP (Canonical Decomposition [2]/Parallel Factor Analysis [3]) decomposition, shifted CP and convolutive CP decomposition will be discussed as well as approaches for the modeling of functional segregation and integration using tensor models with TUCKER2 and PARAFAC2 structure. The talk will finally outline potential areas of neuroscience where tensor modeling may become relevant as well as key challenges modeling neuroimaging data using tensor decomposition [1].

References

- 1 Morten Mørup. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):24–40, 2011
- 2 J. Douglas Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35:283–319, 1970.
- 3 Richard A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA working papers in phonetics*, 16:1–84, 1970.

3.6 Data Fusion using Coupled Matrix and Tensor Factorizations

Evrin Acar (University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license
© Evrim Acar (University of Copenhagen, DK)

Data fusion, i.e., extracting knowledge through the fusion of complementary data sets, is a topic of interest in many fields. For instance, in metabolomics, analytical platforms such as Liquid Chromatography-Mass Spectrometry and Nuclear Magnetic Resonance spectroscopy are used for chemical profiling of biological samples. Measurements from different platforms are capable of detecting different types of chemical compounds with different levels of sensitivity. Jointly analyzing those measurements can provide more accurate characterization and understanding of a physiological/pathological condition. Fusing data from multiple sources has proved useful in various disciplines including metabolomics, neuroscience, social network analysis and signal processing. However, data fusion remains a challenging task since there is a lack of data mining tools that can jointly analyze imperfect (i.e., with missing entries) heterogeneous (i.e., in the form of higher-order tensors and matrices) data sets, and capture underlying shared/unshared structures. In this talk, we discuss the formulation of data fusion as a coupled factorization problem. We give a brief overview of data fusion models based on coupled matrix and tensor factorizations (CMTF), demonstrate the use of various CMTF models with applications from different disciplines and discuss open research problems [1].

References

- 1 E. Acar, R. Bro, and A. K. Smilde. Data fusion in Metabolomics using Coupled Matrix and Tensor Factorizations. *Proceedings of the IEEE*, 103: 1603-1620, 2015.

3.7 Tensors for Representational Learning

Volker Tresp (Siemens AG – München, DE)

License  Creative Commons BY 3.0 Unported license
© Volker Tresp (Siemens AG – München, DE)

Embedding learning, a.k.a. representation learning, has been shown to be able to model large-scale semantic knowledge graphs. A key concept is a mapping of the knowledge graph to a tensor representation whose entries are predicted by models using latent representations of generalized entities. Latent variable models are well suited to deal with the high dimensionality and sparsity of typical knowledge graphs [1]. In recent publications the embedding models were extended to also consider temporal evolutions, temporal patterns and subsymbolic representations. We map embedding models, which were developed purely as solutions to technical problems for modelling temporal knowledge graphs, to various cognitive memory functions, in particular to semantic and concept memory, episodic memory, sensory memory, short-term memory, and working memory.

References

- 1 Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *arXiv preprint arXiv:1503.00759*, 2015

Impulse Talks II: Tensor Computing and Beyond

3.8 Advances in the numerical computation of (coupled) tensor decompositions

Lieven De Lathauwer (KU Leuven, BE)

License  Creative Commons BY 3.0 Unported license
© Lieven De Lathauwer (KU Leuven, BE)

We give a short overview of recent advances in the numerical computation of tensor decompositions and coupled matrix/tensor factorizations. We pay special attention to large-scale aspects. We also highlight some features of Tensorlab, of which version 3.0 has been released in March [1].

References

- 1 Laurent Sorber Marc Van Barel Nico Vervliet, Otto Debals and Lieven De Lathauwer. Tensor lab user guide, release 3.0. <http://www.tensorlab.net/userguide3.pdf>

3.9 Tensors and IoT Computational Issues

Lenore Mullin (University of Albany – SUNY, US)

License  Creative Commons BY 3.0 Unported license
© Lenore Mullin (University of Albany – SUNY, US)

The high-performance computing (HPC) community widely recognizes that the current LINPACK-based benchmark used to establish the TOP500 supercomputer ranking now focuses on the wrong things. It measures near-peak floating-point speed but tolerates relatively low communication performance, but what is now needed to predict application

performance is just the opposite: we need to measure communication performance at all levels of the memory hierarchy, and assume the floating-point hardware is overprovisioned. The solution may reside in the observation that the kernel of LINPACK, matrix-matrix multiplication, is actually a special case of n-dimensional tensor operations. The broader category includes both the Fast Fourier Transform (FFT) and matrix-matrix multiplication that has been optimized by hierarchical blocking to match memory levels [1]. We first unify the generalized task of n-dimensional tensor operations and then show that we can simply dial a different point of the spectrum of such workloads to restore predictive value to the TOP500 ranking on real applications.

References

- 1 Lenore Mullin and James Reynolds. Scalable, portable, verifiable kronecker products on multi-scale computers. In *Constraint Programming and Decision Making*, pages 111–129. Springer, 2014

3.10 Guaranteed Learning Using Tensor Methods

Animashree Anandkumar (University of California – Irvine, US)

License  Creative Commons BY 3.0 Unported license
© Animashree Anandkumar (University of California – Irvine, US)

Unsupervised learning is the challenging problem of making automated discoveries without external supervision. It requires fitting unlabeled data to large-scale latent variable models. Traditional learning approaches such as expectation maximization or variational inference are slow to converge and get stuck in local optima due to non-convexity of the likelihood function. In contrast, we have developed a method of moments approach, based on decomposition of low order moment tensors, which is guaranteed to learn the correct model under mild conditions with (low order) polynomial sample and computational complexity [1].

In practice, tensor methods significantly outperform previous learning approaches, both in training time and model fitting, on a wide range of problems such as document categorization, social network analysis, discovering neuronal cell types, and learning sentence embeddings. Further, we have established that tensor methods are guaranteed to find the globally optimal solution to other challenging non-convex problems such as training multi-layer neural networks and reinforcement learning of partially observable Markov decision processes. These positive results demonstrate that many learning tasks, previously considered intractable, can be solved efficiently under mild and transparent conditions.

References

- 1 Majid Janzamin, Hanie Sedghi, and Anima Anandkumar. Beating the perils of non-convexity: Guaranteed training of neural networks using tensor methods. *CoRR abs/1506.08473*, 2015

3.11 Tensor techniques for the visualization of multidimensional data

Renato Pajarola (Universität Zürich, CH)

License  Creative Commons BY 3.0 Unported license
© Renato Pajarola (Universität Zürich, CH)

We will present the application of tensor approximation methods in the context of interactive visualization of large 3D volume data as well as the processing of visual data in the compressed domain [1].

References

- 1 Renato Pajarola, Susanne K Suter, and Roland Ruiters. Tensor approximation in visualization and computer graphics. 2013

4 Panel discussions

Panel discussions provided us with an additional method of knowledge and experience sharing and exchange of our different perspectives. We distributed panel discussions in two parts over the course of the workshop. In “Part 1: Challenges of Tensor Representations for IoT Data” industrial researchers shared their experience with real-world multidimensional sensor data. In “Part 2: Challenges of Tensor Computing for IoT” panelists shared their view on computational and process challenges with respect to wide spread applicability of tensors in IoT/CPS.

Part I: Challenges of Tensor Representations for IoT data

4.1 Distributed PARAFAC for Large Industrial Internet-scale Dense Tensors with Skewed Modes

Kareem Aggour (General Electric – Niskayuna, US)

License  Creative Commons BY 3.0 Unported license
© Kareem Aggour (General Electric – Niskayuna, US)

Kareem’s work is driven by the need to analyze large Industrial Internet-scale datasets such as those generated by equipment sensors over time. GE is OEM of turbines among other industrial equipment. Typical analysis consists of data from 10-100 turbines, with 60-100 sensors, which deliver per second data. Sensor data is typically noisy. Peculiarities regarding equipment data is that they are both sparse and dense, and very skewed over the time mode. Multimodal concepts, and tensor decomposition lends itself to naturally find relations in this data. Kareem modified the classical PARAFAC algorithm to run in distributed computing environments to decompose large, dense, and skewed 3-way datasets. Distributed PARAFAC is implemented and extensively evaluated on two computing platforms HPC vs. Hadoop. Both scale well with large tensors.

It was mentioned in the audience that there are a few other tools like Splat that focus on sparse and dense data which also enable distributed scaling. Apache Spark also has an implementation for dense tensors.

4.2 Sensors in Power Networks – Representing Network Dynamics over Space and Time

Denis Krompaß (Siemens AG – München, DE)

License  Creative Commons BY 3.0 Unported license
© Denis Krompaß (Siemens AG – München, DE)

Data quality is always challenging, especially in large-scale distributed CPS. There is a wide need for algorithms to handle the preprocessing of data. Use of tensor factorization as feature extractor on sensor data, is promising. A similar challenge is labeling of data. There is not much labeled sensor data – and no easy way to do so reliably – especially not without the constant involvement of a domain expert. Small questions like “how do you normalize the data – per device or per measurement type?” turn out to be bigger design issues. Important issue is that it is not clear to a non-expert where to pay attention in the streams of sensor data. Effective labeling of data is a dialog between domain expert and the data scientist.

Comments from the audience was that latent variables can be exposed in a unique way through tensor decompositions. CP is good for decomposition, good for learning representations, but has problems with numerical analysis. Maybe considering subspaces, as in Tucker, and working with residuals would give a decomposition similar to CP but without the problem of numerical analysis. Additionally, if the decomposed tensor respects the physical model of the system, e.g. power network, that would expose the subspace to focus attention on. Another way could be to guide experts to attach domain knowledge to the data, e.g. which are the relevant/irrelevant areas. Such domain checks could be realized as feedback.

4.3 Loose Semantic Coupling in IoT and the Role of Tensors

Edward Curry for Souleiman Hasan (National University of Ireland – Galway, IE)

License  Creative Commons BY 3.0 Unported license
© Edward Curry for Souleiman Hasan (National University of Ireland – Galway, IE)

Decoupling is a main principle for scalability. Can tensor models provide a principled data representation that is easy to exchange in a decoupled IoT environment? Smart City scenarios deal with data from a lot of distributed sensors, and applications therein are mainly concerned with event processing [1]. Hence, Souleiman and Edward investigated computational semantics by representing the data in vector space and analyzing collocation. The main challenge is how do you find the relevant data on the fly?

References

- 1 Souleiman Hasan and Edward Curry. Thingsonomy: Tackling variety in internet of things events. *IEEE Internet Computing*, 19(2):10–18, 2015

Part II: Challenges of Tensor Computing for IoT

4.4 Key Aspects in Tensor Computing in the Context of IoT

Benoit Meister (Reservoir Labs, Inc. – New York, US)

License  Creative Commons BY 3.0 Unported license
© Benoit Meister (Reservoir Labs, Inc. – New York, US)

Benoit’s working environment mainly applies tensors in cybersecurity applications. The models are shipped with appliances for handling massive throughput, e.g. high-end 100 Gbit/s or 20 Gbit/s. Solutions also include tools that allow the user to select areas for analysis. Security experts use visualization components of the tensor analysis tool. Modeling is done jointly with own domain experts and external security experts. CP is a preferred model because it is easier to communicate, clarify questions for research when structure is interpretable.

The application in IoT scenarios will bring interesting research challenges along, such as faster execution of tensor methods, streaming or low-rank updates for real-time analysis. Efficient distributed data and computation models are required for handling tensor computing in heterogeneous processor architectures with multi-core and embedded systems, operating at low power. Essentially algorithms need a power efficient design. Applying compressed sensing techniques on multidimensional sensor data to reduce computation costs, data movement, and eventually power utilization are promising areas [1].

References

- 1 Nicholas P Carter, Aditya Agrawal, Shekhar Borkar, Romain Cledat, Howard David, Dave Dunning, Joshua Fryman, Ivan Ganey, Roger A Golliver, Rob Knauerhase, et al. Run-nemede: An architecture for ubiquitous high-performance computing. In *High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium on*, pages 198–209. IEEE, 2013

4.5 Modeling of Complex, Man-made Systems

Bülent Yener (Rensselaer Polytechnic Institute – Troy, US)

License  Creative Commons BY 3.0 Unported license
© Bülent Yener (Rensselaer Polytechnic Institute – Troy, US)

Interdisciplinarity is a main reason for huge, repetitive communication overhead in CPS projects. Tensors must be a part of a bigger computational machinery supporting the domain experts. It would be great if we eventually can qualify “it depends on the data”, e.g. through recipes, best practices. For example in IoT/CPS data is always analyzed over time and space. In IoT, prediction (trending) is very important; e.g. anomalies in massive streams of IP traffic data coming from interconnected routers.

4.6 Future Research Directions on Tensor Computing for IoT

Vagelis Papalexakis (Carnegie Mellon University, US)

License © Creative Commons BY 3.0 Unported license
© Vagelis Papalexakis (Carnegie Mellon University, US)

Tensor decompositions are very versatile and powerful tools, ubiquitous in data mining applications. They have been successfully integrating in a rich variety of real-world applications, and due to the fact that they can express and exploit higher order relations in the data, they tend to outperform approaches that ignore such structure. As a result, tensor modeling and analysis in Internet of Things applications, where the data are inherently multi-aspect and diverse, has a great potential [1].

The success that tensors have experienced in data mining during the last few years by no means indicates that all challenges and open problems have been addressed. Quite to the contrary, there exist challenges, such as scaling up to bigger data, modeling space and time, unsupervised model selection, or connections with heterogeneous information networks.

References

- 1 Evangelos E Papalexakis, Christos Faloutsos, and Nicholas D Sidiropoulos. Parcube: Sparse parallelizable tensor decompositions. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 521–536. Springer, 2012

5 Working groups

Break-out Sessions I: Value Proposition of Our Multidisciplinary Research

In the afternoon of the second day, after having shared our diverse perspectives, we took a “map” of our skills and research interest. Especially we were anxious to see how we are going to move from IoT applications of tensor decompositions to running software in decentralizing cyber-physical systems. Below is a snapshot:



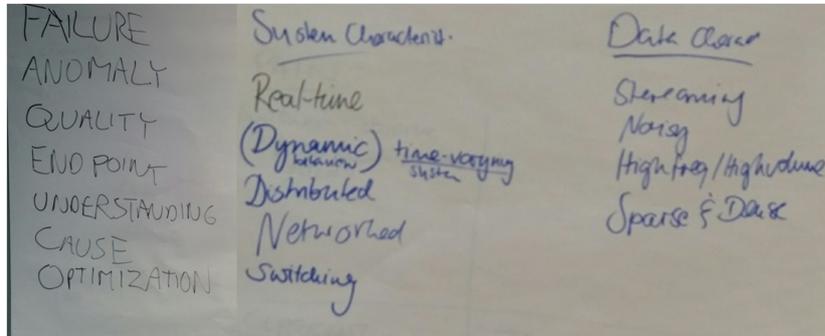
Participants and their area of interest for determining break-out sessions.

Based on this snapshot, we divided into two break-out sessions: Applications and Frameworks.

5.1 Applications of Tensor Decomposition in IoT Scenarios

We discussed typical scenarios in CPS and IoT from domains such as energy, automation, transportation, etc., where operators/users are interested in taking actions based on the

inferred knowledge about faults, anomalies or general events in the system. Especially operators, e.g. in traffic management, or grid management etc. are interested in understanding root cause and ability to optimize system over longer periods as well as in real-time.



IoT/CPS applications, system and data characteristics.

The systems in these domains are real-time, show dynamic behavior, are time-varying, distributed and networked. Usage and operations of these systems include switching, which requires both continuous and discrete variables in representations. The data coming from the sensors embedded into these CPS/IoT systems is streaming, noisy, both high-frequency and high-volume, both sparse and dense.

Many information of interest is conveyed from multiple aspects through different sensors capturing various characteristics of the system. In automation scenarios, when the control loop is closed, system modeling becomes too complicated to model. State space models, such as Kalman filters, are a very natural representation of such closed dynamic systems [1]. State space models are essentially tensorized representations of the physical system.

Novel techniques via tensors, which can handle both inference and modeling, could be applicable for increased interpretability. The above described application, system, data characteristics typically call for a multivariate statistical model of the multi-aspect streaming data. Computational methods in the language of tensors enables the design of provable algorithms, which are otherwise harder to get.

There are also multiple dimensions of challenges. Connectivity for example requires richer tensor models, e.g. hierarchical models. Time dimension requires dynamic processing, e.g. of incoming slices; or active learning of entries, by algorithms that interactively query the user or other data sources to obtain labels. Another approach is the explicit modeling of time or system evolution. The natural structure of tensors to be used both for representation and computationally does simplify the information extraction process. Later arriving data can be updated. For time and connectivity interaction, i.e. change over time that affects connectivity, the whole model can be updated, e.g. via coupled factorization.

On the other hand, the knowledge about the network structure from CPS can be fed into the model as apriori knowledge. Tensor decompositions are more efficient and accurate in representing graphs. An open research question is whether such IoT network graphs are representable through tensor graph networks, if the structure changes dynamically. The connectivity structure of IoT/CPS can help to establish associations between the various modalities capturing data about the system. Tensors can be used to combat curse of dimensionality: tensor methods excel when associations exist in the data. When a few signals dominate, then low rank, i.e. low parametric, representation becomes possible. De-noising of data could also benefit from exploiting the connectivity structure of IoT/CPS for tensor decompositions.

Today in large CPS such as power transmission systems or transportation, logistics systems etc. an important question is how many sensors are needed and where they should be placed. Optimal sensor placement studies need to become dynamic as the underlying systems get more dynamic [2]. Tensors, on the other hand, can uncover hidden structures in sparse data or from compressive resampled data. Monitoring aging devices with sampling only from time to time can already show a trend. Tensor representations are very efficient for such few parameters. Open research questions again arise from the dynamic nature of IoT/CPS in which dynamic states require that the models need to derive state from input/output data.

Many interesting questions arouse when we discussed the need to justify models. Model justification is crucial in CPS and IoT such that system users understand algorithmic inference. The coupling with semantics and links to Multi-agent Systems are a few of the interesting topics to investigate. Modeling of actions in these dynamic systems, especially actions taken based on the intelligence through the algorithmic inference is an open question.

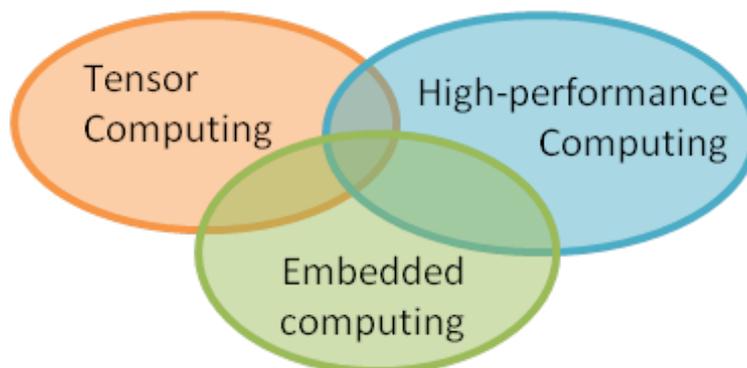
Even in the face of evidence or interpretability of a model, trust into algorithmic inference is a big issue. Domain experts oftentimes underestimate how complex their systems are or will become through increasing digitalization and connectivity. Overreliance on simulations and rule-based systems give a false sense of complexity management. Tensor decompositions are not only interpretable but also well visualizable. New technology for visualization combined with effective visualization methods are open research directions towards creating better understandability of the models and creating trust.

References

- 1 Mohammad Niknazar, Hanna Becker, Bertrand Rivet, Christian Jutten, and Pierre Comon. Robust 3-way tensor decomposition and extended state kalman filtering to extract fetal ecg. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5. IEEE, 2013
- 2 Farrokh Aminifar, Mahmud Fotuhi-Firuzabad, and Amir Safdarian. Optimal pmu placement based on probabilistic cost/benefit analysis. *IEEE Transactions on Power Systems*, 28(1):566–567, 2013

5.2 Frameworks for Tensor Computing in IoT

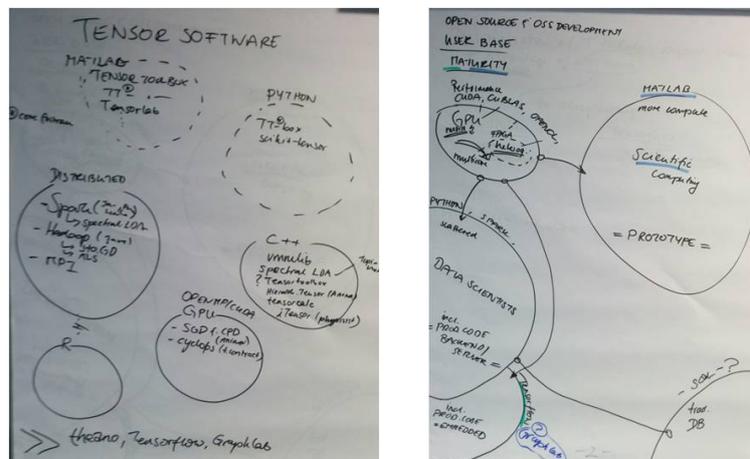
IoT applications will not only require a close collaboration between data scientists, control experts, and domain experts for the application of tensor computing, but also software developers who are proficient in embedded systems, distributed systems, and high performant code.



Rare overlap of skill set requires supporting SW frameworks.

Such a niche overlap as depicted above calls for frameworks abstracting away the complexities of each area to a degree that ultimately IoT application developers should be able to utilize these frameworks with a tolerable learning curve.

In this break-out session we started off by a depiction of currently available tensor software, libraries, and frameworks that are enabling tensor computing. Open source software (OSS) and OSS development is acknowledged as a driving force for increasing user bases and accelerated maturity. At the same time, there are currently two major user groups, data scientists and scientific community/control experts who mostly prefer Python or Matlab respectively. Matlab has a quite mature libraries, such as Tensor Toolbox [1], Tensor Train Toolbox [3], or TensorLab [4]. Python has the NumPy package for scientific computing which has an N-dimensional array object. However, tensor additions such as PyTensor or scikit-tensor [5] have not yet gained community support.

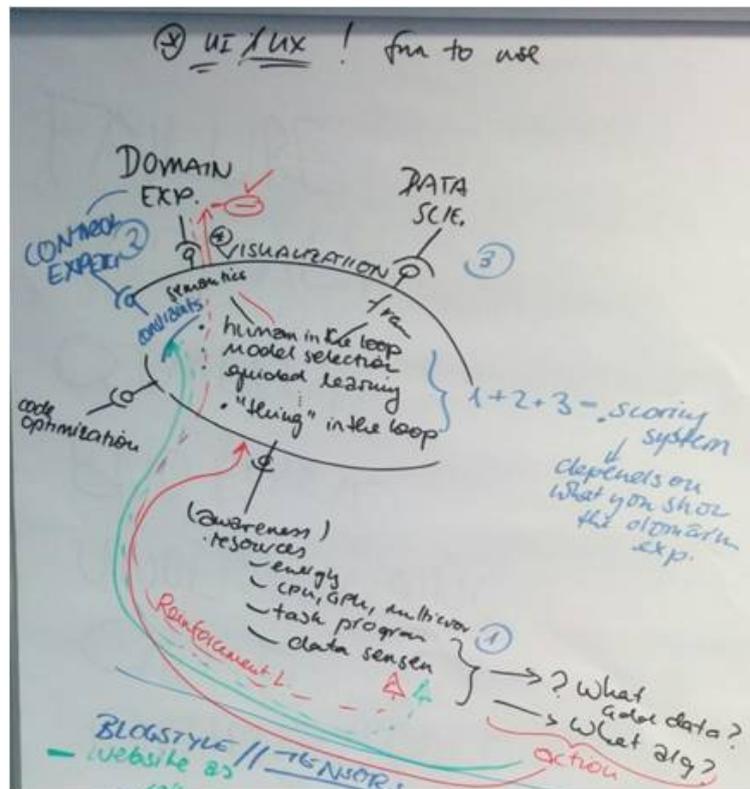


Available SW frameworks and applicability in productive IoT/CPS.

When it comes to production level code for use in IoT applications on massive amounts of streaming data, current frameworks, e.g. Apache Spark [8] or Tensorflow [6] are rather geared towards data scientists with Python as a front end language. Although Spark has SQL interface for traditional DBs [2], the group discussion concluded that traditional relational DBs are less likely to be used in information extraction from IoT. Tensorflow, additionally, enables seamless deployment of production code not only on distributed server clusters but also down to embedded devices.

One important aspect is that currently all these frameworks depend on the same libraries for performant execution of linear algebraic routines, e.g. BLAS, in vendor-specific or -neutral frameworks such as in CUDA or OpenCL, etc. CUDA fine-tunes operations on GPUs of NVIDIA, whereas OpenCL programs execute across heterogeneous platforms including CPUs, GPUs, DSPs, FPGAs or hardware accelerators.

During the course of the discussions we identified that such libraries can be enhanced by taking advantage of tensor algebra. Algebraic analysis can assist in statically solving the optimal partitioning of computations across available processor architectures in an IoT application. Additionally, the Tensorflow framework currently supports deployment on embedded devices [7], but not distributed embedded application, scoring, or update of a trained model. Currently distributed learning is being applied and improved for cluster computing to reduce training times. However, there is no research into distributed embedded application, e.g. retraining, of models.



Potential outline of a Tensors for IoT Framework.

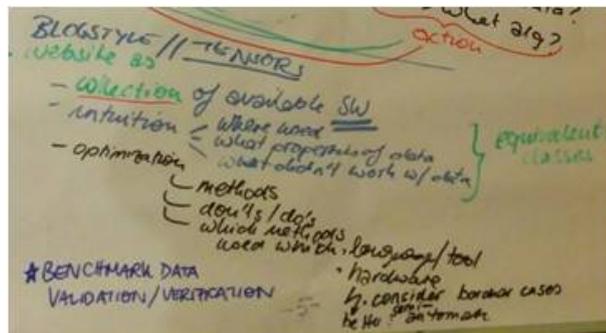
References

- 1 Tamara G. Kolda Brett W. Bader et al. Matlab tensor toolbox version 2.6. <http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html>
- 2 Michael Armbrust, Reynold S Xin, Cheng Lian, Yin Huai, Davies Liu, Joseph K Bradley, Xiangrui Meng, Tomer Kaftan, Michael J Franklin, Ali Ghodsi, et al. Spark sql: Relational data processing in spark. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1383–1394. ACM, 2015
- 3 Ivan Oseledets. Tt (tensor train) toolbox version 2.2.2. <https://github.com/oseledets/TT-Toolbox>
- 4 Laurent Sorber Marc Van Barel Nico Vervliet, Otto Debals and Lieven De Lathauwer. Tensor lab user guide, release 3.0. <http://www.tensorlab.net/userguide3.pdf>
- 5 Maximilian Nickel. scikit-tensor: Python library for multilinear algebra and tensor factorizations. <https://github.com/mnick/scikit-tensor>
- 6 Google. Tensorflow: An open source software library for numerical computation using data flow graphs. <https://www.tensorflow.org/>
- 7 Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zhang. Tensorflow: A system for large-scale machine learning. *CoRR*, abs/1605.08695, 2016
- 8 Apache Spark. Apache spark: A fast and general engine for large-scale data processing. <http://spark.apache.org/>

Break-out Sessions II: Call to Action

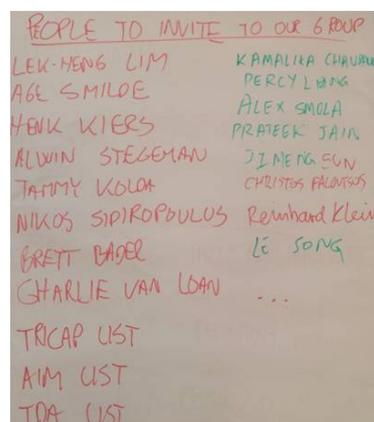
In this break-out session the group concentrated on establishing concrete steps to remove the identified inhibitors towards a multidisciplinary research on tensor decomposition applications in cyber-physical systems:

1. Improve accessibility of tensor methods and knowledge in the CPS/IoT domains:
 - Tensorize Wikipedia – Anima, Vagelis
 - IEEE Proceedings Special Issue for Manifesto – Volker
 - IEEE IoT Paper on Why Tensors – Edward
 - Paper on Multidisciplinary Perspective – Anima, Bülent, Vagelis
 - Paper on Common Notation/Language with 4–10 authors – Vagelis, Rasmus, Taylan, Lieven
2. Create IoT benchmark data repository for a data grand challenge – Sebnem, Edward
 - “30 days of traffic” city benchmark – Edward
 - Satellite/areal imagery – Renato
 - Open smart cities – Edward
 - Seismic data – Bülent, Taylan
3. Connect with tensor operation implementers (in Python, C++,...) – Ivan, Benoit, Sebnem
4. Create a website to link above efforts: community building incl. mailing list, wiki, repository, for defining common problems, data type etc. – Rasmus, Renato



Rough sketch of contents of the website.

5. Invite further experts and practitioners to the community – all



(Incomplete) list of experts to invite to the community.

Additionally, some concrete research interests were formulated based on synergies identified during this workshop

1. Incomplete CP for streaming – Evrim, Vagelis
2. Data Fusion based on moments – Evrim, Anima
3. “Teddy” Tensor decomposition dynamics – Lenore, Sebnem

6 Open Problems

Cyber-physical systems (CPS) or Internet of Things (IoT) describe the now visible trend of increased digitalization and interconnectivity that bridge the physical system and its virtual representation through data from multi-modal sensors embedded into such systems. Network characteristics and dynamics are manifested in the data. Tensor methods are considerably suitable for uncovering such hidden structures, latent variables, in the data if such associations exist. However, there exist challenges, such as scaling up to bigger data, modeling space and time, unsupervised model selection, or connections with heterogeneous information networks.

Model justification is crucial in CPS and IoT such that system users understand algorithmic inference. Modeling of actions in these dynamic systems, especially of how to model the actions taken based on the intelligence gained through the algorithmic inference is an open question. For a better understanding by users, it is of advantage that tensor decompositions are well visualizable. New technology for visualization combined with effective visualization methods are valuable research directions to create traction.

Regarding performance and power efficiency of models and algorithms, we identified that current libraries can be enhanced by taking full advantage of tensor algebra. Algebraic analysis can assist in statically solving the optimal partitioning of computations across available processor architectures in an IoT application. Current frameworks support distributed deployment of tensor computations across a cluster of servers and down to embedded devices. What is missing is distributed embedded application, scoring, or update of a trained model in a streaming real-time manner, which we deem necessary in IoT/CPS applications.

IoT/CPS are dynamic networks. Time and connectivity interactions, i.e. change over time that affects connectivity, require that the whole model can be updated online. One interesting aspect is that the knowledge about network structure from CPS can be fed into the model as apriori knowledge. Tensor decompositions are more efficient and accurate in representing graphs. An open research question is whether network graphs are representable through tensor graph networks, if the structure changes dynamically. Update mechanism and online tensor decompositions need to be researched in order to handle dynamic data. Whether connectivity associations can be taken advantage of in tensor decompositions for better de-noising of data is yet another open question.

7 Outlook

The Dagstuhl Perspectives Workshop was a very suitable format to bring in the multidisciplinary community to exchange knowledge, discuss synergies, and identify potential inhibitors of applications of tensor methods in IoT/CPS. From the many discussions we had, the concrete next steps we identified, one aspect stands out: We want to establish a better foundation

for Tensor Computing in IoT through a knowledge hub website. Such a website has the potential to bring together domain experts and data scientists to exchange best practices. One important pillar of the knowledge hub will be the benchmark IoT data repository that this group wants to harvest.

We believe that IoT/CPS data grand challenges can be an appropriate format to bring domain and control experts together with data scientists effective in applying tensor methods. Over time the community will be able to catalogue and efficiently compare and exchange models, data types, and application types, which we started discussing in this workshop. The hub has the potential to move from knowledge exchange to guided preprocessing of IoT data and model selection. Further more, algorithms embedded in the CPS/IoT could utilize such a hub for active learning and reinforcement learning given sufficient quality content and appropriate interfaces.

In the meantime, we are committed to work on dissemination and seek academic and industrial feedback. Aforementioned activities such as creating and maintaining the knowledge hub, defining IoT data grand challenges to move tensor computing research towards industrial applied research in CPS/IoT will require national and international funding. The potential outcome of more efficient algorithms and better interpretable models for data-driven automation and control of large-scale digitalized systems are of utmost importance to users as well as operators of future electricity, transportation, and manufacturing systems.

Participants

- Evrim Acar
University of Copenhagen, DK
- Kareem Aggour
General Electric – Niskayuna, US
- Animashree Anandkumar
Univ. of California – Irvine, US
- Rasmus Bro
University of Copenhagen, DK
- Ali Taylan Cemgil
Bogaziçi Univ. – Istanbul, TR
- Edward Curry
National University of Ireland – Galway, IE
- Lieven De Lathauwer
KU Leuven, BE
- Hans Hagen
TU Kaiserslautern, DE
- Souleiman Hasan
National University of Ireland – Galway, IE
- Denis Krompaß
Siemens AG – München, DE
- Gerwald Lichtenberg
HAW – Hamburg, DE
- Benoit Meister
Reservoir Labs, Inc. – New York, US
- Lenore Mullin
Univ. of Albany – SUNY, US
- Morten Mørup
Technical Univ. of Denmark – Lyngby, DK
- Axel-Cyrille Ngonga-Ngomo
Universität Leipzig, DE
- Ivan Oseledets
Skoltech – Skolkovo, RU
- Renato Pajarola
Universität Zürich, CH
- Vagelis Papalexakis
Carnegie Mellon University, US
- Christine Preisach
SAP SE – Walldorf, DE
- Achim Rettinger
KIT – Karlsruher Institut für Technologie, DE
- Sebnem Rusitschka
Siemens AG – München, DE
- Volker Tresp
Siemens AG – München, DE
- Bülent Yener
Rensselaer Polytechnic Institute – Troy, US



Report from Dagstuhl Seminar 16161

Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments

Edited by

Elena Cabrio¹, Graeme Hirst², Serena Villata³, and Adam Wyner⁴

- 1 Laboratoire I3S – Sophia Antipolis, FR, elena.cabrio@unice.fr
- 2 University of Toronto, CA, gh@cs.toronto.edu
- 3 Laboratoire I3S – Sophia Antipolis, FR, villata@i3s.unice.fr
- 4 University of Aberdeen, UK, azwyner@abdn.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16161 “Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments”, 17–22 April 2016. The seminar brought together leading researchers from computational linguistics, argumentation theory and cognitive psychology communities to discuss the obtained results and the future challenges of the recently born Argument Mining research area. 40 participants from 14 different countries took part in 7 sessions that included 30 talks, two tutorials, and a hands-on “unshared” task.

Seminar April 17–22, 2016 – <http://www.dagstuhl.de/16161>

1998 ACM Subject Classification I.2.4 Knowledge Representation Formalisms and Methods, I.2.7 Natural Language Processing

Keywords and phrases Argument Mining, Argumentation Theory, Cognitive Science, Computational Linguistics

Digital Object Identifier 10.4230/DagRep.6.4.80

Edited in cooperation with Alexis Palmer

1 Executive Summary

Graeme Hirst

Elena Cabrio

Serena Villata

Adam Wyner

License © Creative Commons BY 3.0 Unported license
© Graeme Hirst, Elena Cabrio, Serena Villata, and Adam Wyner

Philosophers and, in more recent years, theorists working largely within Artificial Intelligence have developed formal computational models of argumentation, how it works, and what makes an argument valid or invalid. This work has made substantial progress in abstract, formal models to represent and reason over complex argumentation structures and inconsistent knowledge bases. Relatively little research, however, has applied these computational models to naturally occurring argumentation in text; nor have Computational Linguistics and Natural Language Processing substantially examined argumentation in text. Moreover, much of the work to date has studied only domain-specific texts and use-cases. Examples include finding the specific claims made in a scientific paper and distinguishing argumentation from narrative in legal texts.



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

Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 80–109

Editors: Elena Cabrio, Graeme Hirst, Serena Villata, and Adam Wyner



DAGSTUHL
REPORTS Dagstuhl Reports

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

But there are many uses and applications for automatic processing of the argumentative aspects of text, such as summarizing the argument of a complex court decision, helping a writer to structure an argument, and processing a large corpus of texts, such as blogs or consumer comments, to find arguments within it. To identify and integrate arguments across a corpus is a very significant problem. To address the issues, solve problems, and build applications, tools must be developed to analyze, aggregate, synthesize, structure, summarize, and reason about arguments in texts. Such tools would enable users to search for particular topics and their justifications, to trace through the argument, and to systematically and formally reason about the relations among arguments. However, to do so requires more linguistic sophistication and newer techniques than currently found in NLP. Moreover, NLP approaches must be connected to computational models of argument. The issues and problems have started to receive attention from both communities; for example, legal documents, on-line debates, product reviews, newspaper articles, court cases, scientific articles, and other kinds of text have all been the subject of recent NLP research on argumentation mining and have been tied to computational models.

Because argumentation is an inherently cross-disciplinary topic involving philosophy, psychology, communications studies, linguistics, and computer science, where different interpretations, analyses, and uses of arguments are proposed and applied, for progress in building NLP tools for argumentation there needs to be progress not only within each domain, but in bridging between these various disciplines, Natural Language Processing, and the computational models. This seminar aimed to help build this bridge by bringing together researchers from different disciplines, with the following goals:

- To understand better the specific kinds of tasks that NLP can carry out in argumentation.
- To establish a set of domain-specific and cross-domain use-cases that will guide the direction of research in the field.
- To understand better how computational argumentation tasks are tied – or not tied – to their specific domains, such as scientific papers, legal argumentation, and political discussions, looking for new cross-domain generalizations.
- To understand better the technical challenges to success in each of these tasks, and to discuss how the challenges can be addressed.
- To develop and explicate specific challenge problems for the integration of argumentation theory and NLP that are beyond the state of the art (but not too much so), and in which success would have the greatest effect on the field.
- To provide prototype solutions that address issues in the integration of NLP and argumentation theory, and to outline follow-on development.
- To propose or provide preliminary solutions to common open challenges in natural language argumentation (among others: argument retrieval in text, argument summarization, identification of semantic relations among arguments), profiting from the cross-fertilization between researchers coming from the research areas of NLP and formal argumentation.

The seminar was held on 17–22 April 2016, with 40 participants from 14 different countries. The event’s seven sessions included 30 talks, two tutorials and a hands-on “unshared” task. The program included several plenary presentations and discussions in smaller working groups. The presentations addressed a variety of topics, as argument mining applied to legal argumentation and to writing support. Collective discussions were arranged for most of these topics, as well as plans for a future interdisciplinary research agenda involving experts from social sciences and psychology.

As a result of the seminar, a number of challenges and open issues have been highlighted:

- At this stage of maturity of the research area, it is difficult to choose good (possibly new) challenges and to define the task(s) to be addressed by automated systems

- Similarly, it is also challenging to precisely define and accomplish annotation task(s) to establish benchmarks and gold standards to test such automated systems
- It is essential to the fruitful development of the research area establish an Interdisciplinary outreach, involving social sciences, psychology, and economics.

Addressing these issues and other questions is now on the agenda of the Argument Mining research community.

2 Table of Contents

Executive Summary

Graeme Hirst, Elena Cabrio, Serena Villata, and Adam Wyner 80

Overview of Talks

Putting Argument Mining to Work: an Experiment in Legal Argument Retrieval
Using the LUIMA Type System and Pipeline
Kevin D. Ashley 85

Expert Stance Graphs for Computational Argumentation
Roy Bar-Haim, Noam Slonim, and Orith Tolego-Ronen 85

Naïve arguments about argument mining
Pietro Baroni 85

Design Reasoning and Design Rationale – an experiment with implications for
argument mining
Floris Bex, Rizkiyanto, and Jan Martijn van der Werf 86

Interfaces to Formal Argumentation
Federico Cerutti 86

Argument Extraction Challenges in a New Web Paradigm
Giorgos Flouris, Antonis Bikakis, Theodore Patkos, and Dimitris Plexousakis 87

On Recognizing Argumentation Schemes in Formal Text Genres
Nancy L. Green 87

Argumentative Writing Support: Structure Identification and Quality Assessment
of Arguments
Iryna Gurevych and Christian Stab 87

Crowdsourced and expert annotations for argument frame discovery
Graeme Hirst 88

Aggregating evidence about the positive and negative effects of treatments
Anthony Hunter and Matthew Williams 90

Introduction to Structured Argumentation
Anthony Hunter 91

Representing and Reasoning about Arguments Mined from Texts and Dialogues
Anthony Hunter, Leila Amgoud, and Philippe Besnard 92

Working on the Argument Pipeline: Through Flow Issues between Natural Language
Argument, Instantiated Arguments, and Argumentation Frameworks
Anthony Hunter, Adam Wyner, and Tom van Engers 92

Temporal Argument Mining for Writing Assistance
Diane J. Litman 93

Locating and Extracting Key Components of Argumentation from Scholarly Sci-
entific Writing
Robert Mercer 93

Argumentation Mining in Online Interactions: Opportunities and Challenges
Smaranda Muresan 94

Argument Strength Scoring <i>Vincent Ng</i>	96
Clause types in argumentative texts <i>Alexis M. Palmer and Maria Becker</i>	96
Joint prediction in MST-style discourse parsing for argumentation mining <i>Andreas Peldszus and Manfred Stede</i>	97
Strategical Argumentative Agent for Human Persuasion <i>Ariel Rosenfeld</i>	98
Towards Knowledge-Driven Argument Mining <i>Patrick Saint-Dizier</i>	99
Medication safety as a use case for argumentation mining <i>Jodi Schneider and Richard D. Boyce</i>	99
Social Media Argumentation Mining: The Quest for Deliberateness in Raucousness <i>Jan Šnajder</i>	100
Assessing Argument Relevance at Web Scale <i>Benno Stein and Henning Wachsmuth</i>	100
Towards Relation-based Argumentation Mining? <i>Francesca Toni</i>	102
The Need for Annotated Corpora from Legal Documents, and for (Human) Protocols for Creating Them: The Attribution Problem <i>Vern R. Walker</i>	102
Working groups	
A Pilot Study in Mining Argumentation Frameworks from Online Debates <i>Federico Cerutti, Alexis M. Palmer, Ariel Rosenfeld, Jan Šnajder, and Francesca Toni</i>	103
An Unshared Untask for Argumentation Mining <i>Ivan Habernal and Adam Wyner</i>	103
Participants	109

3 Overview of Talks

3.1 Putting Argument Mining to Work: an Experiment in Legal Argument Retrieval Using the LUIMA Type System and Pipeline

Kevin D. Ashley (University of Pittsburgh, US)

License © Creative Commons BY 3.0 Unported license
© Kevin D. Ashley

This paper highlights some results from an experiment demonstrating the feasibility of legal argument retrieval of case decisions in a particular legal domain. The LUIMA program annotates legal texts in terms of the roles some of the sentences play in a legal argument. It uses the annotations of argument-related information to re-rank cases returned by conventional legal information retrieval methods in order to improve the relevance of the top-ranked cases to users' queries. The experiment assessed the effectiveness of the re-ranking empirically and objectively, demonstrating how argument mining can be put to work.

3.2 Expert Stance Graphs for Computational Argumentation

Roy Bar-Haim, Noam Slonim, and Orith Tolego-Ronen

License © Creative Commons BY 3.0 Unported license
© Roy Bar-Haim, Noam Slonim, and Orith Tolego-Ronen
Main reference Will be presented in the 3rd Workshop on Argument Mining ACL 2016, Berlin.
URL <http://argmining2016.arg.tech/>

We describe the construction of an Expert Stance Graph, a novel, large-scale knowledge resource that encodes the stance of more than 100,000 experts towards a variety of controversial topics. We suggest that this graph may be valuable for various fundamental tasks in computational argumentation. Experts and topics in our graph are Wikipedia entries. Both automatic and semi-automatic methods for building the graph are explored, and manual assessment validates the high accuracy of the resulting graph.

3.3 Naïve arguments about argument mining

Pietro Baroni (University of Brescia, IT)

License © Creative Commons BY 3.0 Unported license
© Pietro Baroni

This discussion talk presents some questions about argument mining from an outsider perspective, with the goal of discussing some basic alternatives in approaching this important research problem, pointing out some aspects that may be worth clarifying to a newcomer to the field, and providing hints about what other research fields in the area of computational argumentation can offer. In particular the talk touches the following issues: holistic vs. restricted approaches, simple abstract vs. detailed argumentation models, domain dependence vs. context dependence, annotated corpora vs. paradigmatic examples, mining vs. generating arguments, and a priori vs. discovered ontologies.

3.4 Design Reasoning and Design Rationale – an experiment with implications for argument mining

Floris Bex (Utrecht University, NL), Rizkiyanto, and Jan Martijn van der Werf

License © Creative Commons BY 3.0 Unported license
© Floris Bex, Rizkiyanto, and Jan Martijn van der Werf

In system and software design, we are faced with a combination of design reasoning, the process of analysing issues, options, pros and cons, and design rationale, the product of this process in the form of explicit issues, options, pros and cons. In our research we propose a basic card game aimed at improving the design reasoning process. This card game was used in an experiment, and the resulting design deliberations were transcribed and coded, where special care was taken to annotate the different design rationale elements posed by the participants. This, we can investigate the links between design reasoning phases and design rationale elements. In future research, the aim is to use process mining to analyse the design reasoning and, ultimately, try to (semi-)automatically mine design rationale from design reasoning discussions.

3.5 Interfaces to Formal Argumentation

Federico Cerutti (Cardiff University, GB)

License © Creative Commons BY 3.0 Unported license
© Federico Cerutti

Joint work of Federico Cerutti, Nava Tintarev, and Nir Oren
Main reference F. Cerutti, N. Tintarev, N. Oren, “Formal Arguments, Preferences, and Natural Language Interfaces to Humans: an Empirical Evaluation”, in Proc. of the 21st European Conference on Artificial Intelligence (ECAI’14), Frontiers in Artificial Intelligence and Applications, Vol. 263, pp. 207–212, IOS Press, 2014.
URL <http://dx.doi.org/10.3233/978-1-61499-419-0-207>

Like other systems for automatic reasoning, argumentation approaches can suffer from “opacity.” We explored one of the few mixed approaches explaining, in natural language, the structure of arguments to ensure an understanding of their acceptability status. In particular, we summarised the results described in [1], in which we assessed, by means of an experiment, the claim that computational models of argumentation provide support for complex decision making activities in part due to the close alignment between their semantics and human intuition. Results show a correspondence between the acceptability of arguments by human subjects and the justification status prescribed by the formal theory in the majority of the cases. However, post-hoc analyses show that there are some deviations. This seems to suggest that there is the need for some effort for making formal argumentation process more transparent.

References

- 1 Federico Cerutti, Nava Tintarev, and Nir Oren. *Formal Arguments, Preferences, and Natural Language Interfaces to Humans: an Empirical Evaluation*. In *21st European Conference on Artificial Intelligence*, pages 207–212, 2014.

3.6 Argument Extraction Challenges in a New Web Paradigm

Giorgos Flouris (FORTH – Heraklion, GR), Antonis Bikakis, Theodore Patkos, and Dimitris Plexousakis

License © Creative Commons BY 3.0 Unported license
© Giorgos Flouris, Antonis Bikakis, Theodore Patkos, and Dimitris Plexousakis

The exchange of comments, opinions and arguments in blogs, social media, commercial websites or wikis is transforming the Web into a modern agora, a virtual place where all types of debates take place. This wealth of information remains unexploited: the purely textual form of online arguments leaves limited room for automated processing and the available methodologies of computational argumentation focus on logical arguments, failing to properly model online debates. We envision a formal, machine-interpretable representation of debates that would enable the discovery, tracking, retrieval, combination, interrelation, extraction and visualization of the vast variety of viewpoints that already exist on the Web, in a way that goes beyond simple keyword-based processing. This paper describes this vision and the related challenges, focusing on challenges related to argument extraction.

3.7 On Recognizing Argumentation Schemes in Formal Text Genres

Nancy L. Green (University of North Carolina – Greensboro, US)

License © Creative Commons BY 3.0 Unported license
© Nancy L. Green

Argumentation mining research should address the challenge of recognition of argumentation schemes in formal text genres such as scientific articles. This paper argues that identification of argumentation schemes differs from identification of other aspects of discourse such as argumentative zones and coherence relations. Argumentation schemes can be defined at a level of abstraction applicable across the natural sciences. There are useful applications of automatic argumentation scheme recognition. However, it is likely that inference-based techniques will be required.

3.8 Argumentative Writing Support: Structure Identification and Quality Assessment of Arguments

Iryna Gurevych (TU Darmstadt, DE) and Christian Stab

License © Creative Commons BY 3.0 Unported license
© Iryna Gurevych and Christian Stab
Main reference C. Stab, I. Gurevych, “Parsing Argumentation Structures in Persuasive Essays”, arXiv:1604.07370v2 [cs.CL], 2016.
URL <https://arxiv.org/abs/1604.07370v2>

Argumentation is an omnipresent daily routine. We engage argumentation not only for making decisions or convincing an audience but also for drawing widely accepted conclusions and inferring novel knowledge. Good argumentation skills are crucial to learning itself. Consequently, argumentation constitutes an important part of education programs. With the emergence of the *Common Core Standard*, argumentative writing receives increasing attention. However, many students are still underprepared in writing well-reasoned arguments since

teachers are not able to provide sufficient writing assignments in view of increasing class sizes and the load for reviewing argumentative texts. In addition, current *Intelligent Writing Systems* (IWS) are limited to feedback about spelling, grammar, mechanics, and discourse structures, and there is no system that provides feedback about written arguments. Novel *Natural Language Processing* (NLP) methods that analyze written arguments and pinpoint the weak points in argumentative discourse could bridge this gap.

In this talk, we highlighted the recent research projects on *Computational Argumentation* (CA) at the UKP-Lab. In particular, we provided an overview of the methods developed in the context of *Argumentative Writing Support* (AWS). We highlighted the results of an annotation study on argumentation structures, introduced an argumentation structure annotated corpus of persuasive essays, and presented a novel end-to-end *argumentation structure parser* for extracting micro-level argumentation structures. We introduced the architecture of the system and outlined the evaluation results. In addition, we introduced two novel tasks and our experimental results on *quality assessment* of natural language arguments. First, we presented our work on identifying myside bias in persuasive essays. Second, we presented an approach for identifying insufficiently supported arguments.

3.9 Crowdsourced and expert annotations for argument frame discovery

Graeme Hirst (University of Toronto, CA)

License © Creative Commons BY 3.0 Unported license
© Graeme Hirst

Joint work of Naderi, Nona; Hirst, Graeme

Introduction

Theoretical perspectives on framing and frames are diverse, but these theories converge in their conceptualization of framing as a communication process to present an object or an issue. According to Entman (1993):

Framing involves selection and salience. To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described.

In parliamentary discourse, politicians use framing to highlight some aspect of an issue when they expound their beliefs and ideas through argumentation. For example, a politician speaking about immigration, regardless of which side of the issue they are on, might frame it as one of economics, or of multiculturalism, or of social justice.

The long-term goal of our work is to build computational models for the automatic analysis of issue-framing in argumentative political speech. In this paper, we address the short-term goal of annotating issue-framing in argumentative political speech.

Crowd-driven annotation?

We experimented with crowd-driven annotation of the Parliamentary debates to create training and test data. Such an annotation task requires the crowd members to have time to read through the speeches, to have knowledge of the issue at hand, and to have skill at identifying the correct frame. Our experiments used text from the debates of the Canadian

Parliament on two recent issues: The proposed federal gun registry (1995) and gay marriage (instituted in Canada in 2005 but subsequently re-debated in Parliament in 2005–6). We selected paragraphs from Parliamentary debates on these topics and manually (intuitively) derived a set of frames and topics for each side of issue. We then created a CrowdFlower task: Annotate each paragraph with its frames and topics. Multiple selections were permitted as was “none of the above”. For the gun-registry topic, we had 1200 paragraphs and identified 22 frames. They included: gun ownership is a privilege, not a right; people need guns for protection; and a gun registry would reduce crime. For gay marriage, we had 659 paragraphs and 13 frames. They included: gay marriage is unnatural as it can’t produce children; and marriage is a human right.

The results demonstrated that this task was too difficult and was not possible for crowd members. After 16 days, we had only 285 judgements on the gun-registry texts, of which just 17 were trusted, and similar results on the gay-marriage texts. We tried making the task easier by reducing the number of frames to nine, but this made little difference. Moreover, there was little agreement between the crowd members and some expert annotations that we made.

Experiments on frame identification

We therefore turned to a smaller, simpler dataset as training data: the ComArg corpus of online debates in English, annotated with a set of these frames (Boltužić & Šnajder 2014). Prior manual analysis of online debates yielded set of ‘standard’ arguments for specific topics – in effect, frames for the topics – one of which was gay marriage for which there were seven standard arguments (three pro and four con). We used these as training data for identifying frames in Canadian parliamentary debates on the same topic.

But we still needed annotated Parliamentary texts for evaluation, so instead of crowd-sourcing, we asked our expert colleagues to act as annotators for 400 selected paragraph-length texts (two annotators each) and 136 selected sentences (three annotators each). We used only texts for which at least two annotators were in agreement, leaving us with 366 paragraph-length texts and 121 sentences.

To classify each Parliamentary text by frame, we used its textual similarity to the Boltužić & Šnajder arguments. We tried three different vector-based representations:

- word2vec embeddings (300 dimensions) (Mikolov et al. 2013), summing the vectors for the words to get a vector for the complete sentence or paragraph.
- Syntax-based word embeddings (300 dimensions) (Wang et al. 2015), again summed.
- Skip-thought sentence vectors (4800 dimensions) (Kiros et al. 2015).

For measurements of textual similarity to the statement of the argument we tried both the cosine between the vectors and concatenation of absolute difference between the vectors with their component-wise product (“p&d”). We used these measurements as features in a multiclass support-vector machine. In addition, we also tried using known stance as a possible additional feature. Our baselines were the majority class and using a simple bag-of-words similarity with $tf \cdot idf$ weights.

We obtained better accuracy with the p&d similarity measure, but there was no consistent pattern in the results with respect to representation. For the paragraph-length texts, the majority-class baseline was an accuracy of 53.3%, the bag-of-words baseline was 71.0% accuracy, and the best accuracy achieved by a combination of our methods was 75.4%; most combinations gave results at or well below baseline accuracy. For the single sentences, the baselines were 33.0% and 52.8% respectively, and our best accuracy was 73.5%; most combinations gave results lower than that and only a few points above baseline. (Complete

tables of results are given by Naderi and Hirst (2016).) Although, as noted, there was no consistent pattern in the results, we observed that the syntactically informed representation never gave good results, which was a surprise as one would have expected it to be superior at least to the purely lexical word2vec representation.

Conclusion

In this preliminary study, we examined annotation and automatic recognition of frames in political argumentative discourse. It is very preliminary work, using a noisy, simplistic representation of frames as text. We found that most of the information used by the classifier is lexical and that syntax is surprisingly unhelpful. We also found that annotation of the texts was too difficult for crowdsourcing, which severely limits the size of frame-annotated corpora that we can build. We will explore semi-supervised or unsupervised approaches as an alternative.

Acknowledgements. This work is supported by the Natural Sciences and Engineering Research Council of Canada and by the Social Sciences and Humanities Research Council. We thank Patricia Araujo Thaine, Krish Perumal, and Sara Scharf for their contributions to the annotation of parliamentary statements, and Tong Wang for sharing the syntactic embeddings. We also thank Tong Wang and Ryan Kiros for fruitful discussions.

References

- 1 Boltužić, F., Šnajder, J. (2014). Back up your stance: Recognizing arguments in online discussions. *Proceedings of the First Workshop on Argumentation Mining*, pp. 49–58.
- 2 Entman, R.M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication* 43(4), 51–58 (1993)
- 3 Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R.S., Torralba, A., Urtasun, R., Fidler, S. (2015). Skip-thought vectors. arXiv preprint arXiv:1506.06726.
- 4 Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- 5 Naderi, Nona and Hirst, Graeme (2016). Argumentation mining in parliamentary discourse. To appear. (Available on request from the authors.)
- 6 Wang, Tong; Mohamed, Abdel-rahman; Hirst, Graeme (2015). Learning lexical embeddings with syntactic and lexicographic knowledge. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 458–463.

3.10 Aggregating evidence about the positive and negative effects of treatments

Anthony Hunter (University College London, GB) and Matthew Williams

License © Creative Commons BY 3.0 Unported license
© Anthony Hunter and Matthew Williams

Main reference A. Hunter, M. Williams, “Aggregating evidence about the positive and negative effects of treatments”, *Artificial Intelligence in Medicine*, 56(3):173–190, 2012.

URL <http://dx.doi.org/10.1016/j.artmed.2012.09.004>

Objectives: Evidence-based decision making is becoming increasingly important in health-care. Much valuable evidence is in the form of the results from clinical trials that compare

the relative merits of treatments. In this paper, we present a new framework for representing and synthesizing knowledge from clinical trials involving multiple outcome indicators.

Method: The framework generates and evaluates arguments for claiming that one treatment is superior, or equivalent, to another based on the available evidence. Evidence comes from randomized clinical trials, systematic reviews, meta-analyses, network analyses, etc. Preference criteria over arguments are used that are based on the outcome indicators, and the magnitude of those outcome indicators, in the evidence. Meta-arguments attacks arguments that are based on weaker evidence.

Results: We evaluated the framework with respect to the aggregation of evidence undertaken in three published clinical guidelines that involve 56 items of evidence and 16 treatments. For each of the three guidelines, the treatment we identified as being superior using our method is a recommended treatment in the corresponding guideline.

Conclusions: The framework offers a formal approach to aggregating clinical evidence, taking into account subjective criteria such as preferences over outcome indicators. In the evaluation, the aggregations obtained showed a good correspondence with published clinical guidelines. Furthermore, preliminary computational studies indicate that the approach is viable for the size of evidence tables normally encountered in practice.

3.11 Introduction to Structured Argumentation

Anthony Hunter (University College London, GB)

License  Creative Commons BY 3.0 Unported license
© Anthony Hunter

Joint work of Besnard, Philippe; Hunter, Anthony

Main reference Ph. Besnard, A. Hunter, “Constructing Argument Graphs with Deductive Arguments: A Tutorial”, *Argument and Computation*, 5(1):5–30, 2014.

URL <http://dx.doi.org/10.1080/19462166.2013.869765>

In this tutorial of structural argumentation, we focus on deductive argumentation. A deductive argument is a pair where the first item is a set of premises, the second item is a claim, and the premises entail the claim. This can be formalized by assuming a logical language for the premises and the claim, and logical entailment (or consequence relation) for showing that the claim follows from the premises. Examples of logics that can be used include classical logic, modal logic, description logic, temporal logic, and conditional logic. A counterargument for an argument A is an argument B where the claim of B contradicts the premises of A . Different choices of logic, and different choices for the precise definitions of argument and counterargument, give us a range of possibilities for formalizing deductive argumentation. Further options are available to us for choosing the arguments and counterarguments we put into an argument graph. If we are to construct an argument graph based on the arguments that can be constructed from a knowledgebase, then we can be exhaustive in including all arguments and counterarguments that can be constructed from the knowledgebase. But there are other options available to us. We consider some of the possibilities in this review [1].

References

- 1 Ph Besnard and A Hunter (2014) Constructing Argument Graphs with Deductive Arguments: A Tutorial, *Argument and Computation*, 5(1):5–30

3.12 Representing and Reasoning about Arguments Mined from Texts and Dialogues

Anthony Hunter (University College London, GB), Leila Amgoud, and Philippe Besnard

License © Creative Commons BY 3.0 Unported license

© Anthony Hunter, Leila Amgoud, and Philippe Besnard

Main reference L. Amgoud, Ph. Besnard, A. Hunter, “Representing and Reasoning About Arguments Mined from Texts and Dialogues,” in Proc. of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU’15), LNCS, Vol. 9161, pp. 60–71, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-20807-7_6

Main reference L. Amgoud, Ph. Besnard, A. Hunter, “Logical Representation and Analysis for RC-Arguments”, . in Proc. of the 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’15), pp. 104–110, IEEE CS, 2015.

URL <http://dx.doi.org/10.1109/ICTAI.2015.28>

This talk presents a target language for representing arguments mined from natural language. The key features are the connection between possible reasons and possible claims and recursive embedding of such connections. Given a base of these arguments and counterarguments mined from texts or dialogues, we want be able combine them, deconstruct them, and to analyse them (for instance to check whether the set is inconsistent). To address these needs, we propose a formal language for representing reasons and claims, and a framework for inferencing with the arguments and counterarguments in this formal language [1, 2].

References

- 1 Leila Amgoud, Philippe Besnard, Anthony Hunter: Representing and Reasoning About Arguments Mined from Texts and Dialogues. ECSQARU 2015: 60–71
- 2 Leila Amgoud, Philippe Besnard, Anthony Hunter: Logical Representation and Analysis for RC-Arguments. ICTAI 2015: 104–110

3.13 Working on the Argument Pipeline: Through Flow Issues between Natural Language Argument, Instantiated Arguments, and Argumentation Frameworks

Anthony Hunter (University College London, GB), Adam Wyner, and Tom van Engers

License © Creative Commons BY 3.0 Unported license

© Anthony Hunter, Adam Wyner, and Tom van Engers

In many domains of public discourse such as arguments about public policy, there is an abundance of knowledge to store, query, and reason with. To use this knowledge, we must address two key general problems: first, the problem of the knowledge acquisition bottleneck between forms in which the knowledge is usually expressed, e.g. natural language, and forms which can be automatically processed; second, reasoning with the uncertainties and inconsistencies of the knowledge. Given such complexities, it is labour and knowledge intensive to conduct policy consultations, where participants contribute statements to the policy discourse. Yet, from such a consultation, we want to derive policy positions, where each position is a set of consistent statements, but where positions may be mutually inconsistent. To address these problems and support policy-making consultations, we consider recent automated techniques in natural language processing, instantiating arguments, and reasoning with the arguments in argumentation frameworks. We discuss application and “bridge” issues between these techniques, outlining a pipeline of technologies whereby: expressions in a controlled natural language are parsed and translated into a logic (a literals and rules

knowledge base), from which we generate instantiated arguments and their relationships using a logic-based formalism (an argument knowledge base), which is then input to an implemented argumentation framework that calculates extensions of arguments (an argument extensions knowledge base), and finally, we extract consistent sets of expressions (policy positions). The paper reports progress towards reasoning with web-based, distributed, collaborative, incomplete, and inconsistent knowledge bases expressed in natural language.

References

- 1 Adam Wyner, Tom van Engers, and Anthony Hunter. Working on the Argument Pipeline: Through Flow Issues between Natural Language Argument, Instantiated Arguments, and Argumentation Frameworks. *Argument & Computation*, to appear, 2016.

3.14 Temporal Argument Mining for Writing Assistance

Diane J. Litman (University of Pittsburgh, US)

License © Creative Commons BY 3.0 Unported license
© Diane J. Litman

Joint work of Zhang, Fan; Litman, Diane K.

The written arguments of students are educational data that can be automatically mined for purposes of student instruction and assessment. While prior work has focused on argument mining within a single version of a text, our work focuses on temporal argument mining across a text and its revisions. This paper will illustrate some of the opportunities and challenges in temporal argument mining. I will briefly summarize how we are using natural processing to develop a temporal argument mining system, and how our system in turn is being embedded in an educational technology for providing writing assistance.

3.15 Locating and Extracting Key Components of Argumentation from Scholarly Scientific Writing

Robert Mercer (University of Western Ontario – London, CA)

License © Creative Commons BY 3.0 Unported license
© Robert Mercer

Joint work of Graves, Heather; Graves, Roger; Ibn Faiz, Syeed; Houngbo, Hospice; Ansari, Shifta; Alliheedi, Mohammed; DiMarco, Chrysanne; Mercer, Robert

Mining the components of argumentation suggested by the Toulmin model from the text of scholarly scientific writings has been one research focus. Two such efforts are highlighted: examining titles as a source for the argumentative claim in experimental biomedical articles, and extracting higher order relations between two biomedical relations that correspond closely to warrants. The talk also presents an automated method to produce a sliver standard corpus for IMRaD sentence classification. Other sources of information pertaining to arguments are briefly introduced: (1) data in scientific writing is presented in tables and figures requiring non-linguistic methods to mine this information, (2) scientific arguments are not local to a single research article, so information from other articles could enhance the understanding of the argument in the article and place it in its broader scientific context, and (3) having a representation of how science is performed and then written about in the form of arguments could be beneficial.

References

- 1 Shifta Ansari, Robert E. Mercer, and Peter Rogan. 2013. Automated phenotype-genotype table understanding. In *Contemporary Challenges and Solutions in Applied Artificial Intelligence*, volume 489 of *Studies in Computational Intelligence*, pages 47–52. Springer.
- 2 Heather Graves, Roger Graves, Robert E. Mercer, and Mahzereen Akter. 2014. Titles that announce argumentative claims in biomedical research articles. In *Proceedings of the First Workshop on Argumentation Mining*, pages 98–99.
- 3 Hospice Houngho and Robert E. Mercer. 2014. An automated method to build a corpus of rhetorically-classified sentences in biomedical texts. In *Proceedings of the First Workshop on Argumentation Mining*, pages 19–23.
- 4 Syeed Ibn Faiz and Robert E. Mercer. 2013. Identifying explicit discourse connectives in text. In *Proceedings of the 26th Canadian Conference on Artificial Intelligence (AI'2013)*, pages 64–76.
- 5 Syeed Ibn Faiz and Robert E. Mercer. 2014. Extracting higher order relations from biomedical text. In *Proceedings of the First Workshop on Argumentation Mining*, pages 100–101.
- 6 Mohammad Syeed Ibn Faiz. 2012. Discovering higher order relations from biomedical text. Master's thesis, University of Western Ontario, London, ON, Canada.
- 7 Barbara White. 2009. *Annotating a corpus of biomedical research texts: Two models of rhetorical analysis*. Ph.D. thesis, The University of Western Ontario, Canada.

3.16 Argumentation Mining in Online Interactions: Opportunities and Challenges

Smaranda Muresan (Columbia University – New York, US)

License © Creative Commons BY 3.0 Unported license
© Smaranda Muresan

Joint work of Aakhus, Mark; Ghosh, Debanjan; Muresan, Smaranda; Wacholder, Nina
Main reference D. Ghosh, S. Muresan, N. Wacholder, M. Aakhus, M. Mitsui, “Analyzing Argumentative Discourse Units in Online Interactions”, in Proc. of the First Workshop on Argumentation Mining at ACL, pp. 39–48, ACL, 2014.

URL <http://acl2014.org/acl2014/W14-21/pdf/W14-2106.pdf>

Argument mining of online interactions is in its infancy. One reason is the lack of annotated corpora in this genre. Another reason is that the coding of text as argument often misses how argument is an interactive, social process of reasoning. To make progress, we need to develop a principled and scalable way of determining which portions of texts are argumentative and what is the nature of argumentation. In this talk, I highlighted our approach to argumentation mining in online interactions that places a premium on identifying what is *targeted* and how it is *called out* (Ghosh et al., 2014; Wacholder et al., 2014; Aakhus, Muresan and Wacholder, 2013), and then I discussed some of the opportunities and challenges we face in this area.

Our approach defines an argumentative structure that contains the most basic components of interactive argumentation (i.e., CallOut, Target and Argumentative Relation) as well as finer-grained characteristics of CallOuts (e.g., Stance and Rationale components) and of Argumentative Relations (e.g., type Agree/Disagree/Other) (Ghosh et al., 2014; Wacholder et al., 2014). Our annotation study followed four desiderata: 1) Start with a coarse-grained argumentative structure (e.g., CallOut, Target, Argumentative Relations), and move to more fine-grained argumentative structures; 2) Code not just argument components (Target and CallOut) but also Argumentative Relations between these components; 3) Identify boundary points of argumentative discourse units (ADUs), which in principle can be of any length; and 4) Design an annotation granularity/scope that enables innovation in annotation by

combining traditional and crowd-sourcing practices involving expert and novice annotators. We annotated a dataset of blog posts and their comments. Computationally, we tackled one specific problem: classifying the type of Argumentative Relation between a CallOut and a Target using local models (Ghosh et al., 2014).

The talk concluded with the discussion of several open issues:

- *Segmentation Step.* In our annotation study, the segmentation subtask (identifying argumentative discourse units), proved to be challenging since annotators were free to choose text spans of any length. First, our results show variation in the number of ADUs (e.g., CallOuts) identified by the expert annotators. The consistent variation among coders indicated that, beyond any potential training issues, annotators could be characterized as “lumpers”, who treat a single long segment of text as one ADU and “splitters”, who treat it as two (or more) shorter ADUs (Wacholder et al., 2014). Second, the segmentation variability meant that measurement of IAA had to account for fuzzy boundaries. Third, for developing computational models the segmentation step will be particularly challenging.
- *Computational Models.* Argument Relations are often long-distance and implicit. Most approaches to argumentation mining rely on local models. Recently, Peldszus and Stede (2015) proposed a global model that jointly predicts different aspects of the argument structure. A challenge will be to develop such global models based on discourse-parsing for online interactions, where we do not only have inter-sentence and intra-sentence relations, but also inter-turns relation.
- *Varied Datasets of Argumentative Texts.* As highlighted also by the Unshared Untask at the Dagstuhl seminar, there is a growing need for a common repository of different types of argumentative texts. For online interactions, our corpus consists of blog posts and their comments. Another source of online interactions is Reddit. A particularly interesting subreddit for argumentation mining is ChangeMyView, where posters are “people who have an opinion on something but accept that they may be wrong or want help changing their view.” If a user is able to change someone else’s view, they are awarded a delta point. Recently a large collection of ChangeMyView data has been released to the research community (Tan et al., 2016). An opportunity will be for working groups that focus on various aspects of argumentation to annotate the same datasets.

References

- 1 Ghosh, D., Muresan, S. , Wacholder, N. , Aakhus, M. and Mitsui, M. (2014). Analyzing argumentative discourse units in online interactions. *Proceedings of the First Workshop on Argumentation Mining*, 2014.
- 2 Aakhus, M., Muresan, S. and Wacholder, N (2013). Integrating natural language processing and pragmatic argumentation theories for argumentation support. *Proceedings of the 10th International Conference of the Ontario Society for the Study of Argumentation (OSSA 10): Virtues of Argumentation*, D. Mohammed and M. Lewiski (Eds.), 2013.
- 3 Wacholder, N., Muresan, S., Ghosh, D. and Aakhus, M. (2014). Annotating Multiparty Discourse: Challenges for Agreement Metrics. *Proceedings of LAW VIII – The 8th Linguistic Annotation Workshop*, 2014.
- 4 Peldszus, Andreas and Stede, Manfred (2015). Joint prediction in MST-style discourse parsing for argumentation mining. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015.
- 5 Tan, C., Niculae, V, Danescu-Niculescu-Mizil, C., and Lee, L. (2016). Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. *In Proceedings of the 25th International Conference on World Wide Web (WWW '16).*, 2016.

3.17 Argument Strength Scoring

Vincent Ng (*University of Texas at Dallas, US*)

License © Creative Commons BY 3.0 Unported license
© Vincent Ng

Joint work of Persing, Isaac; Ng, Vincent

Main reference I. Persing, V. Ng, “Modeling Argument Strength in Student Essays,” in Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int’l Joint Conf. on Natural Language Processing (Volume 1: Long Papers), pp. 543–552, ACL, 2015.

URL <http://aclweb.org/anthology/P/P15/P15-1053.pdf>

While recent years have seen a surge of interest in automated essay grading, including work on grading essays with respect to particular dimensions such as prompt adherence, coherence, and technical quality, there has been relatively little work on grading the essay dimension of argument strength. Argument strength, which refers to the persuasiveness of the argument an essay makes for its thesis, is arguably the most important aspect of argumentative essays. In this talk, I will introduce a corpus of argumentative student essays annotated with argument strength scores and propose a supervised, feature-rich approach to automatically scoring the essays along this dimension. I will conclude this talk with a discussion of the major challenges associated with argument strength scoring and argumentation mining in student essays.

3.18 Clause types in argumentative texts

Alexis M. Palmer (*Universität Heidelberg, DE*) and Maria Becker

License © Creative Commons BY 3.0 Unported license
© Alexis M. Palmer and Maria Becker

Main reference M. Becker, A. Palmer, A. Frank, “Argumentative texts and clause types”, in Proc. of the 3rd Workshop on Argument Mining, pp. 21–30, ACL 2016.

URL <http://aclweb.org/anthology/W/W16/W16-2803.pdf>

This work is built on the theoretical framework of discourse mode theory, in which types of text passages are linked to linguistic characteristics of the clauses which compose the passages. [3] identifies five discourse modes: Narrative, Description, Report, Information, and Argument/Commentary. One way in which modes differ from one another is in their characteristic distributions of clause types. Taking the argumentative microtext corpus [2] as a set of prototypical argumentative text passages, we apply Smith’s typology of clause types – know as situation entity (SE) types. The aim is to better understand what types of situations (states, events, generics, etc.) are most prevalent in argumentative texts and, further, to link this level of analysis to the argumentation graphs provided with the microtext corpus. The annotation project is ongoing, but preliminary analysis confirms that argumentative texts do in fact look different from non-argumentative texts with respect to SE types. This result suggests the potential for using systems for automatic prediction of SE types [1] to support computational mining of arguments from texts.

References

- 1 Annemarie Friedrich and Alexis Palmer and Manfred Pinkal. *Situation entity types: automatic classification of clause-level aspect*. In Proc. of ACL 2016. Berlin, Germany, 2016
- 2 Andreas Peldszus and Manfred Stede. *An annotated corpus of argumentative microtexts*. In Proceedings of the First European Conference on Argumentation: Argumentation and Reasoned Action. Lisbon, Portugal, 2015
- 3 Carlota S. Smith. *Modes of discourse: The local structure of texts*. Cambridge University Press, 2003

3.19 Joint prediction in MST-style discourse paring for argumentation mining

Andreas Peldszus (Universität Potsdam, DE) and Manfred Stede (Universität Potsdam, DE)

License  Creative Commons BY 3.0 Unported license
© Andreas Peldszus and Manfred Stede

We introduce two datasets for argumentation mining: a selection of pro & contra newspaper commentaries taken from the German daily “Tagesspiegel”, and a set of user-generated “microtexts”. These have been produced in response to a trigger question such as “BLA”, and consist of about 5 sentences, all of which are relevant for the argumentation. The microtext corpus consists of 112 texts, which have been translated from German to English. [5]

Our annotation scheme [3] used for both corpora builds on the proposals of [1] and constructs a full argument structure for a text: A central claim is (possibly recursively) backed by supporting statements. Following the inherently dialogical nature of argumentation, there may also be potential objections (by an “opponent”) and accompanying counter-objections (by the “proponent”). Our experiments show that these can be reliably annotated by experts ($\kappa=0.83$). In addition, we conducted experiments with lightly-trained students and employed clustering methods to identify reliable annotators, who also reached good agreement [2].

For automatic analysis of microtext argumentation, we developed an approach that divides structure prediction into four classification subtasks: central claim; statement’s perspective as proponent or opponent; function as support or attack, and attachment between statements. We train individual local models for these, and then combine their predictions in a data structure we call “evidence graph”, as it combines the individual contributions and allows for computing the globally preferred structure that respects certain well-formedness constraints. This process is implemented as minimal-spanning tree computation. It is guaranteed to yield complete and sound structures, and for the four subtasks, it gives significant improvements over the results of the local models [4].

Finally, we present an extension of the microtext corpus with two additional layers of discourse structure annotation: Rhetorical Structure Theory and Segmented Discourse Representation Theory. To achieve comparability, we first harmonized the segmentation decisions of the discourse layers and the argumentation layer. This allowed us to map all layers into a common dependency format based on identical discourse segments. To illustrate the potential of correlating the layers, we show how relations from RST correspond to those in the argumentation structure. In future work, on the one hand it is possible to systematically study the relationship between RST and SDRT. On the other hand, it will be fruitful to explore how argumentation mining can benefit from the presence of discourse structure information. [6]

References

- 1 Freeman, J. B.: *Argument Structure: Representation and Theory*. Argumentation Library (18). Springer, 2011.
- 2 Andreas Peldszus: Towards segment-based recognition of argumentation structure in short texts. *Proceedings of the First Workshop on Argumentation Mining*. ACL 2014. Baltimore, Maryland. pp. 88–97
- 3 Andreas Peldszus and Manfred Stede: From Argument Diagrams to Argumentation Mining in Texts: A Survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)* Vol. 7, No. 1, pp. 1–31. 2013

- 4 Andreas Peldszus and Manfred Stede: Joint prediction in MST-style discourse parsing for argumentation mining. Proc. of EMNLP 2015, Lisbon.
- 5 Andreas Peldszus, Manfred Stede. An annotated corpus of argumentative microtexts. To appear in: First European Conference on Argumentation: Argumentation and Reasoned Action, Portugal, Lisbon, June 2015.
- 6 Manfred Stede, Stergos Afantenos, Andreas Peldszus, Nicholas Asher and Jérémy Perret: Parallel Discourse Annotations on a Corpus of Short Texts. Proc. of LREC 2016. Portoroz, Slovenia

3.20 Strategical Argumentative Agent for Human Persuasion

Ariel Rosenfeld (Bar-Ilan University – Ramat Gan, IL)

License  Creative Commons BY 3.0 Unported license
© Ariel Rosenfeld

Main reference Strategical Argumentative Agent for Human Persuasion, 2016, Submitted.

Automated agents should be able to persuade people in the same way people persuade each other – via dialogs. Today, automated persuasion modeling and research use unnatural assumptions regarding the persuasive interaction which creates doubt regarding their applicability for real-world deployment with people. In this work we present a novel methodology for persuading people through argumentative dialogs. Our methodology combines theoretical argumentation modeling, machine learning and Markovian optimization techniques that together result in an innovative agent named SPA. Two extensive field experiments, with more than 100 human subjects, show that SPA is able to persuade people significantly more often than a baseline agent and no worse than people are able to persuade each other. This study is part of our ongoing effort to investigate the connections and challenges between Argumentation Theory and people [2, 1]. We hope that the encouraging results shown in this work (and in previous ones) will inspire other researchers in the field to investigate other argumentation-based methods in human experiments. We believe that bridging the gap between formal argumentation and human argumentation is essential for making argumentation practical for a wider range of applications.

References

- 1 A. Rosenfeld and S. Kraus. Providing arguments in discussions based on the prediction of human argumentative behavior. In *AAAI*, pages 1320–1327, 2015.
- 2 Ariel Rosenfeld and Sarit Kraus. 2014. Argumentation theory in the field: An empirical study of fundamental notions. In *Frontiers and Connections between Argumentation Theory and Natural Language Processing*.

3.21 Towards Knowledge-Driven Argument Mining

Patrick Saint-Dizier (CNRS-Paul Sabatier University – Toulouse, FR)

License © Creative Commons BY 3.0 Unported license
© Patrick Saint-Dizier

Main reference J., Pustejovsky, “The Generative Lexicon”, MIT Press, 1995.

Main reference P. Saint-Dizier, “Processing natural language arguments with the TextCoop platform”, *Journal of Argumentation and Computation*, 3(1), 2012.

Given a controversial issue, argument mining from texts in natural language is extremely challenging: besides linguistic aspects, domain knowledge is often required together with appropriate forms of inferences to identify arguments. This contribution explores the types of knowledge and reasoning schemes that are required and how they can be paired with language resources to accurately mine arguments. We show, via corpus analysis, that the Generative Lexicon (GL) structure, enhanced in different manners and associated with inferences and language patterns, is a relevant approach to capture the typical concepts found in arguments.

3.22 Medication safety as a use case for argumentation mining

Jodi Schneider (University of Pittsburgh, US) and Richard D. Boyce

License © Creative Commons BY 3.0 Unported license
© Jodi Schneider and Richard D. Boyce

URL <http://www.slideshare.net/jodischneider/medication-safety-as-a-use-case-for-argumentation-mining-dagstuhl-seminar-16161-2016-0419>

We present a use case for argumentation mining, from biomedical informatics, specifically from medication safety. Tens of thousands of preventable medical errors occur in the U.S. each year, due to limitations in the information available to clinicians. Current knowledge sources about potential drug-drug interactions (PDDIs) often fail to provide essential management recommendations and differ significantly in their coverage, accuracy, and agreement. The Drug Interaction Knowledge Base Project (Boyce, 2006-present; dikb.org) is addressing this problem.

Our current work is using knowledge representations and human annotation in order to represent clinically-relevant claims and evidence. Our data model incorporates an existing argumentation-focused ontology, the Micropublications Ontology. Further, to describe more specific information, such as the types of studies that allow inference of a particular type of claim, we are developing an evidence-focused ontology called DIDEO–Drug-drug Interaction and Drug-drug Interaction Evidence Ontology. On the curation side, we will describe how our research team is hand-extracting knowledge claims and evidence from the primary research literature, case reports, and FDA-approved drug labels for 65 drugs.

We think that medication safety could be an important domain for applying automatic argumentation mining in the future. In discussions at Dagstuhl, we would like to investigate how current argumentation mining techniques might be used to scale up this work. We can also discuss possible implications for representing evidence from other biomedical domains.

3.23 Social Media Argumentation Mining: The Quest for Deliberateness in Raucousness

Jan Šnajder (University of Zagreb, HR)

License  Creative Commons BY 3.0 Unported license

© Jan Šnajder

Joint work of Šnajder, Jan; Boltužić, Filip

Argumentation mining from social media content has attracted increasing attention. The task is both challenging and rewarding. The informal nature of user-generated content makes the task dauntingly difficult. On the other hand, the insights that could be gained by a large-scale analysis of social media argumentation make it a very worthwhile task. In this position paper I discuss the motivation for social media argumentation mining, as well as the tasks and challenges involved.

3.24 Assessing Argument Relevance at Web Scale

Benno Stein (Bauhaus-Universität Weimar, DE) and Henning Wachsmuth

License  Creative Commons BY 3.0 Unported license

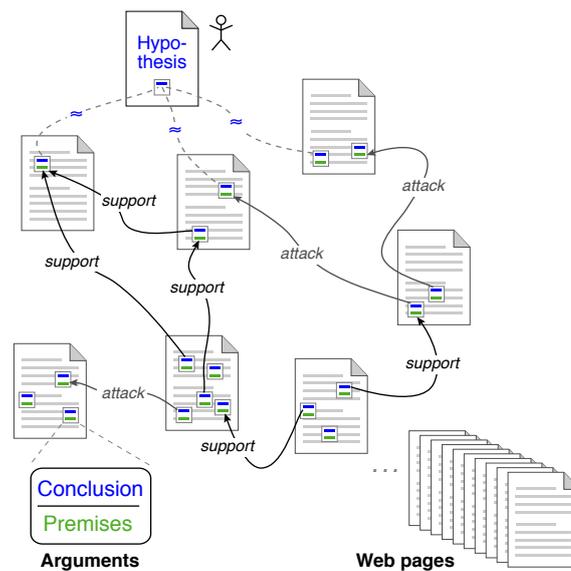
© Benno Stein and Henning Wachsmuth

Joint work of Stein, Benno; Wachsmuth, Henning

The information needs of users will focus more and more on arguments that can be found pro and con a queried hypothesis [3]. As a consequence, future information systems, above all web search engines, are expected to provide justifications for the results they return in response to user queries [7].

Accordingly, argument mining has become an emerging research topic, also being studied for the web [1]. Argument mining identifies the units of arguments (i.e., premises and conclusions) in natural language texts and it classifies their relations, but it does not clarify which arguments are *relevant* for a given hypothesis. First approaches to assess argument strength or similar exist [6, 2]. However, they hardly account for the fundamental problem that argument relevance is essentially subjective.

In our work, we envision the structural and hence objective assessment of argument relevance at web scale. To draw a clear line between existing work and the missing building blocks, we presume that technologies are available which can (1) robustly mine argument units from web pages and (2) decide if two units mean the same—or maybe the opposite. Based hereupon, we model all argument units and relations found on the web in an argument graph. We devise an adaptation of the famous PageRank algorithm [5], which recursively processes the argument graph to compute a score for each argument unit. From these scores, the relevance of arguments can be derived. The depicted figure below sketches a small argument graph. If we interpret the hypothesis of a user as a conclusion—as shown—all arguments with that conclusion may be relevant for the user's information need.



Originally, PageRank aims to measure the objective relevance of a web page based on all other pages that link to that page. Similarly, for an argument, we measure its relevance based on all other arguments that make use of its conclusion. Thereby, we separate conclusion relevance from the soundness of the inference an argument makes to arrive at its conclusion. In analogy to the supportive nature of web links, we restrict our “PageRank for argument relevance” to supporting argument units here. However, a variant based on attack relations is conceivable as well, covering ideas from [4] for the web then.

Practically, the construction of a reliable argument graph raises complex challenges of processing natural language texts. Also, the adapted PageRank brings up new questions, e.g., whether and how to balance support and attack. Following our approach, however, these questions can be addressed stepwise to bring argument relevance to search engines, starting from the technologies of today.

References

- 1 Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein. *Cross-Domain Mining of Argumentative Text through Distant Supervision*. In Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics, pages 1395–1404, 2016.
- 2 Liora Braunstain, Oren Kurland, David Carmel, Idan Szpektor, and Anna Shtok. *Supporting Human Answers for Advice-Seeking Questions in CQA Sites*. In Proceedings of the 38th European Conference on IR Research, pages 129–141, 2016.
- 3 Elena Cabrio and Serena Villata. *Natural Language Arguments: A Combined Approach*. In Proceedings of the 20th European Conference on Artificial Intelligence, pages 205–210, 2012.
- 4 Phan Minh Dung. *On the Acceptability of Arguments and its Fundamental Role in Non-monotonic Reasoning, Logic Programming and n-Person Games*. Artificial Intelligence, 77(2):321–357, 1995.
- 5 Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report 1999-66, Stanford InfoLab, 1999.
- 6 Isaac Persing and Vincent Ng. *Modeling Argument Strength in Student Essays*. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and

the 7th International Joint Conference on Natural Language Processing: Volume 1: Long Papers, pages 543–552, 2015.

- 7 Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. *Show Me Your Evidence – An Automatic Method for Context Dependent Evidence Detection*. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 440–450, 2015.

3.25 Towards Relation-based Argumentation Mining?

Francesca Toni (Imperial College London, GB)

License © Creative Commons BY 3.0 Unported license
© Francesca Toni

Joint work of Lucas Carstens

Main reference L. Carstens, F. Toni, “Towards relation based argumentation mining”, in Proc. of the 2nd Workshop on Argumentation Mining, affiliated with NAACL 2015, pp. 29–34, ACL, 2015.

URL <http://aclweb.org/anthology/W/W15/W15-0504.pdf>

In this talk I overviewed foundations, tools and applications of Structured argumentation (ABA) and Abstract Argumentation (AA) for rule-based arguments as well as Bipolar argumentation and Quantitative Argumentation Debates (QuADs), see [1] for an overview. These frameworks can be supported by and support the mining of attack/support/neither relations amongst arguments (e.g. as in [2]). Moreover, I have discussed the following questions: is the use of quantitative measures of strength of arguments, as proposed e.g. in QuADs, a good way to assess the dialectical strength of mined arguments or the doodness of argument mining? Can argumentation help argument mining?

References

- 1 Lucas Carstens, Xiuyi Fan, Yang Gao, and Francesca Toni. An overview of argumentation frameworks for decision support. In Madalina Croitoru, Pierre Marquis, Sebastian Rudolph, and Gem Stapleton, editors, Graph Structures for Knowledge Representation and Reasoning – 4th International Workshop, GKR 2015, Buenos Aires, Argentina, July 25, 2015, Revised Selected Papers, volume 9501 of Lecture Notes in Computer Science, pages 32–49. Springer, 2015.
- 2 Lucas Carstens and Francesca Toni. Towards relation based argumentation mining. In The 2nd Workshop on Argumentation Mining, affiliated with NAACL 2015, 2015.

3.26 The Need for Annotated Corpora from Legal Documents, and for (Human) Protocols for Creating Them: The Attribution Problem

Vern R. Walker (Hofstra University – Hempstead, US)

License © Creative Commons BY 3.0 Unported license
© Vern R. Walker

This presentation argues that in order to make progress today in automating argumentation mining from legal documents, we have a critical need for two things. First, we need a sufficient supply of manually annotated corpora, as well as theoretical and experimental evidence that those annotated data are accurate. Second, we need protocols for effectively training people to perform the tasks and sub-tasks required to create those annotations. Such protocols are necessary not only for a team approach to annotation and for quality assurance of the

finished annotations, but also for developing and testing software to assist humans in the process of annotation. Drawing upon the latest work at Hofstra University’s Law, Logic and Technology Research Laboratory in New York, the paper offers an extended example from the problem of annotating attribution relations, as an illustration of why obtaining consistent and accurate annotations in law is extremely difficult, and of why protocols are necessary. Attribution is the problem of determining which actor believes, asserts, or relies upon the truth of a proposition as a premise or a conclusion of an argument. The paper illustrates that in applying argumentation mining to legal documents, annotating attribution relations correctly is a critical task.

4 Working groups

4.1 A Pilot Study in Mining Argumentation Frameworks from Online Debates

Federico Cerutti (Cardiff University, GB), Alexis M. Palmer (Universität Heidelberg, DE), Ariel Rosenfeld (Bar-Ilan University – Ramat Gan, IL), Jan Šnajder (University of Zagreb, HR), and Francesca Toni (Imperial College London, GB)

License © Creative Commons BY 3.0 Unported license
© Federico Cerutti, Alexis M. Palmer, Ariel Rosenfeld, Jan Šnajder,
and Francesca Toni

We describe a pilot study mapping an online debate onto several types of Argumentation Frameworks, as understood in the AI field of Computational Argumentation. The pilot study aims to explore the richness of online debates *and* of Computational Argumentation methods and techniques. Additionally we consider the potential benefits of connecting the output of Argument Mining and the tools offered by Computational Argumentation, in particular algorithms in existing Argumentation Frameworks for determining the dialectical acceptability or strength of arguments. The mapping makes use of an intermediate graphical representation of the debate, manually generated by human annotators.

4.2 An Unshared Untask for Argumentation Mining

Ivan Habernal (TU Darmstadt, DE) and Adam Wyner

License © Creative Commons BY 3.0 Unported license
© Ivan Habernal and Adam Wyner

Introduction

In this extended abstract, we outline the *Unshared Untask* at the Argument Mining seminar at Dagstuhl. Argument mining (also known as “Argumentation mining”) is a recent challenge in corpus-based discourse processing that applies a certain argumentation theory to model and automatically analyze the data at hand. Interest in Argument Mining has rapidly increased over the past few years, as demonstrated by a number of events (The BiCi seminar Frontiers and Connections between Argumentation Theory and Natural Language Processing in 2014; Workshops on Argumentation Mining at ACL 2014, NAACL 2015, and ACL 2016; the Dagstuhl Seminar on Debating Technologies in 2015; and the Dagstuhl seminar reported

here on Natural Language Argumentation, to name a few). Given the wide the range of different perspectives and approaches within the community, it is now very important to consolidate the view and shape the further development of research on Argument Mining.

Shared tasks have become a major driver in boosting research in many NLP fields.¹ The availability of shared annotated data, clear evaluation criteria, and visible competing systems allow for fair, exact comparison between systems and overall progress tracking, which in turn fosters future research. However, argument mining, as an evolving research field, suffers not only from the lack of large data sets but also from the absence of a unified perspective on what tasks the systems should fulfill and how the systems should be evaluated. The tasks and evaluation measures require agreement, which is particularly problematic given the variety of argumentation models, argumentative genres and registers, granularity (e.g., micro-argumentation and macro-argumentation), dimensions of argument (logos, pathos, ethos), and the overall social context of persuasion.

The concept of a so-called “unshared untask” is an alternative to shared tasks. In an unshared untask, neither a clearly defined problem to be solved nor quantitative performance measures are given. Instead, participants are given only a variety of raw unannotated data and an open-ended prompt. The goals are to explore possible tasks, try to provide a rigorous definition, propose an annotation methodology, and define evaluation metrics, among others. This type of activity has been successful in several areas, such as PoliInformatics (a broad group of computer scientists, political scientists, economists, and communications scholars) [2] and clinical psychology and NLP [1]. Some venues also run shared and unshared tasks in parallel [3]. The nature of an unshared untask is mainly exploratory and can eventually lead to a deeper understanding of the matter and to laying down foundations for a future standard shared task.

The remainder of this extended abstract outlines the process, data, and observations about this activity.

Process

For this Dagstuhl seminar, we collected several diverse, relatively small samples as the basis for individual analysis and group discussion. We did not set any agenda or framework around analysis or outcome; rather, the individuals and groups were free to contribute what and as they thought appropriate. The objective was to generate a broad spectrum of ideas about the issues and challenges as well as the approaches and techniques.

In order to keep the ‘idea pots stirring’ in some organized fashion, we had created two ‘rings’ of three groups, and each ring worked with different datasets. Each group in a ring processed all the datasets in its ring over the course of three meetings.

For the closing panel discussion, we prepared and elicited critical questions to assess multiple aspects of the proposed tasks with respect to the corresponding data. This session summarized main observations and revealed possible future directions for the community, as will be presented later in this paper.

Data

We prepared the following data samples in a ‘raw’ form; the source texts were transformed into a plain-text format with only minimal structure kept (such as IDs of comments in Web

¹ See the various tracks in TREC <http://trec.nist.gov/>

discussions to properly identify which post a user is responding to). All visual cues (images, different font sizes) or non-related texts (HTML boiler-plate) were removed.

Scientific Paper Reviews contains three reviews and author responses from the NIPS 2013-2014 conferences.² Our motivation for this data was that (1) reviews of scientific articles should contain a strictly factual and precise argumentation and (2) this register has not yet been tackled in argumentation mining, to the best of our knowledge.

Amazon Camera Reviews is a collection of 16 product reviews from Amazon written by users. While reviews have been heavily exploited in the sentiment analysis field, there have been only few works on argumentation mining in this genre.

Twitter is a collection of 50+50 random Tweets related to ‘Brexit’ collected over the period of two weeks in April 2016 using hashtags such as #strongerin or #leaveeu. The collection was cleaned in terms of removing all re-tweets, tweets containing images or links to articles, and keeping only tweets without mentions of other Twitter accounts. This filtering step should have ensured that the tweets were written by the users ‘from scratch’ with the intention to support either staying in the EU or leaving it.

Policy Making is a collection of material derived from policy-making consultations that are carried out by the European Union on the topic of copyright for academic access. So far as we know, there have been no works on argumentation mining in this genre. The outcomes of the consultations are used to develop directives. The dataset is made up of responses by stakeholders to particular policy-making queries circulated by the commission. Thus, different stakeholders, e.g. companies and academic libraries, may have very different and contrastive views. The participants were provided with the *Green Paper: Copyright in the Knowledge Economy*, three sample responses, and two questions to focus attention on (questions 4 and 5).³

Grab Bag

Along with the four data sources presented above, we provided a ‘grab bag’ with a more diverse collection of data which were either solicited from the seminar participants or provided as additional datasets by the organizers. The following list introduces data that were discussed by at least one group in the break-out sessions.

Debate Portals is a subset of data (variant A) provided for the Unshared Task for the 3rd Workshop on Argument Mining co-located with ACL 2016 in Berlin.⁴ Samples originate from a two-sided debate portal createdebate.com.

Opening and closing speeches from Oxford-style debates is a subset of data (variant B) provided for the Unshared Task for the 3rd Workshop on Argument Mining co-located with ACL 2016.

News Editorials with two articles about the currency in the Scottish Independence Referendum.

Persuasive Essay Revisions are two text files which represent the first and second drafts of a persuasive essay (related to educationally oriented argument mining for writing support).⁵

² <http://papers.nips.cc/>

³ On http://ec.europa.eu/internal_market/copyright/copyright-info/index_en.htm#maincontentSec2 see *Green Paper* and *The Replies to the Public Consultation* for ENPA, FAEP, and British Library.

⁴ <https://github.com/UKPLab/argmin2016-unshared-task/>

⁵ Contributed by Diane Litman

Medial Article is an excerpt from a scientific article containing an abstract, author summary, introduction, and discussion.⁶

Observations and discussion

This section attempts to summarize the main observations and possible future tasks for each particular dataset collected across groups and rings. During the final plenary discussion, we performed a poll in order to assess each data type with respect to following critical questions.

Critical Questions

Appropriateness Is the dataset “argumentative” enough to be worth investigating further?

Impact Is the task “attractive” enough to gain more visibility for our community?

Novelty Has the task already been tackled in computational argumentation or Computational Linguistics?

Visibility Would a shared task on this data attract researchers also from outside our community?

Comprehensibility What amount of domain knowledge is required to understand the data?

Reproducibility Do we have enough resources to be annotated and made freely available?

Scalability Is the task feasible for annotation by crowds?

Relativity Is the representation relative to individuals or groups?

Computability Can the representation/annotation/analysis be formalized?

Community Relevance Is the corpus and its analysis relevant to audience / community X?

Scientific Paper Reviews

It was pointed out by many that scientific reviews and author responses do follow some kind of latent structure. One particular direction of analysis thus can focus on identifying structural parts that deal with *clarity*, *originality*, and similar aspects of a scientific piece of work.

Proposed downstream tasks can tackle the primary outcome of a review, namely recognizing whether the submitted paper is good or bad; this coarse grained analysis, however, partly overlaps with sentiment analysis reviews. For that, the final marks of a review can be used as a proxy for annotated data. A more fine-grained task that goes deeper into the argumentation structures might thus involve *inconsistencies* between the the reviewer arguments (comments) and the final related marks given. Moreover, identifying reasons for *rejection*, *acceptance*, or *revisions* were identified as meaningful application outcomes.

Taking into account the dialog between the author’s response and the original reviews, there are two options for analysis. The first use case is to analyze authors’ rebuttals to the reviewers’ comments in order to reveal which arguments were addressed. The second use case is to compare reviews from several reviewers and find contradictions or inconsistencies.

Overall, the intention in argumentative analysis of reviews (and optionally authors’ responses) is to support scientific reviewing. The plenary discussion tackled not only the possible argument mining tasks and applications in scientific reviews, but also incorporating more elaborated reviewing strategies (such as providing reviewers with argument templates, or even some sort of claim and premise form). However, these issues tackle fundamental

⁶ Contributed by Nancy Green

reviewing policies in the particular communities and are beyond the scope of the discussion here.

The plenum agreed that this dataset is very *appropriate* for further studying and a possible domain for a shared task, and is *novel* with a possible big *impact*. It would also help gain more *visibility* of computational argumentation. However, the level of expertise required for annotating the data is high, which might hinder *scalability*.

Amazon Camera Reviews

While product reviews have been a widespread data source for example in the sentiment analysis field, the major concern of the task participants was whether this genre is relevant to computational argumentation at all. Since fine-grained sentiment analysis (e.g., aspect-based sentiment) can already deal with a good level of information extraction from reviews and has reached maturity to be deployed in business, the added value of analyzing arguments in reviews remains an open question. Therefore the obvious tasks, such as extracting reasons for buyer's decisions seem not to be attractive enough, as the reasoning in reviews is usually based on listing pros and cons. Usefulness of reviews and its relation to argumentation was discussed during the plenary session but no clear consensus was reached.

Twitter

The main message taken from analyzing the Brexit-related tweets was that Twitter users try to be as cryptic as possible and rather than arguing, they tend to show off. One big obstacle is the need of background knowledge for the given topic in order to detect the stance or assess relevance of the presented arguments. Only occasionally, full-fledged arguments were presented. However, some potential tasks using Twitter data might include stance detection (already introduced at SemEval-2016), detecting a scheme of an argument, or mining relevant controversial sub-topics (such as the degree of control, costs of campaigns, etc., in case of Brexit). Overall, Twitter has not been found to be strongly appropriate as a shared task for argument mining in the near future, however, its specifics (short messages, language, instant nature) make it a very challenging resource.

Policy making

While the argumentative and legal nature of the consultation make this dataset an interesting resource for studying argumentation, the task participants spent most of the effort on understanding the context, argumentative intention and decoding the main message. It was observed that the arguments are often about values as are other policy-making procedures. The nature of this dataset was found to be too difficult for the shared task, as it requires lots of background information to make sense of the subtle lines of argumentation. Therefore a possible high-level task could involve intention recognition. Overall, the dataset was found to be very appropriate for future argumentation mining research, if broken down into smaller feasible sub-tasks.

Future Work

As previous section highlighted, the discussion about the unshared untask highlighted criteria to identify relevant corpora rather than analysis tasks, though the identification of the corpora was largely determined by efforts to carry out some analysis. Thus, it remains for future work to:

- identify a range of tasks to exercise on a given corpus.
- develop ways to distinguish argumentative from non-argumentative texts as well as different genres within corpora of argumentative texts.
- offer evaluation metrics for multi-dimensional aspects of argumentation mining.
- incorporate context, point-of-view, background, and implicit/enthymatic information into argumentation mining.

References

- 1 Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 31–39, Denver, Colorado, 2015. Association for Computational Linguistics.
- 2 Noah A Smith, Claire Cardie, Anne Washington, and John Wilkerson. Overview of the 2014 NLP Unshared Task in PoliInformatics. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 5–7, Baltimore, MD, USA, 2014. Association for Computational Linguistics.
- 3 Marcos Zampieri, Liling Tan, Nikola Ljubešić, Jörg Tiedemann, and Preslav Nakov. Overview of the DSL Shared Task 2015. In *Proceedings of the Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects*, pages 1–9, Hissar, Bulgaria, 2015. Association for Computational Linguistics.

Participants

- Laura Alonso Alemany
National Univ. – Córdoba, AR
- Michal Araszkiwicz
Univ. Jagiellonski – Krakow, PL
- Kevin D. Ashley
University of Pittsburgh, US
- Pietro Baroni
University of Brescia, IT
- Floris Bex
Utrecht University, NL
- Katarzyna Budzynska
Polish Academy of Sciences –
Warsaw, PL
- Elena Cabrio
Laboratoire I3S – Sophia
Antipolis, FR
- Claire Cardie
Cornell University, US
- Federico Cerutti
Cardiff University, GB
- Giorgos Flouris
FORTH – Heraklion, GR
- Nancy L. Green
University of North Carolina –
Greensboro, US
- Iryna Gurevych
TU Darmstadt, DE
- Ivan Habernal
TU Darmstadt, DE
- Graeme Hirst
University of Toronto, CA
- Eduard H. Hovy
Carnegie Mellon University –
Pittsburgh, US
- Anthony Hunter
University College London, GB
- Diane J. Litman
University of Pittsburgh, US
- Bernardo Magnini
Bruno Kessler Foundation –
Trento, IT
- Robert Mercer
University of Western Ontario –
London, CA
- Marie-Francine Moens
KU Leuven, BE
- Smaranda Muresan
Columbia Univ. – New York, US
- Vincent Ng
University of Texas at Dallas, US
- Sebastian Padó
Universität Stuttgart, DE
- Fabio Paglieri
CNR – Rome, IT
- Alexis M. Palmer
Universität Heidelberg, DE
- Andreas Peldszus
Universität Potsdam, DE
- Ariel Rosenfeld
Bar-Ilan University –
Ramat Gan, IL
- Patrick Saint-Dizier
CNRS-Paul Sabatier University –
Toulouse, FR
- Jodi Schneider
University of Pittsburgh, US
- Edwin Simpson
TU Darmstadt, DE
- Noam Slonim
IBM – Haifa, IL
- Jan Šnajder
University of Zagreb, HR
- Manfred Stede
Universität Potsdam, DE
- Benno Stein
Bauhaus-Universität Weimar, DE
- Simone Teufel
University of Cambridge, GB
- Francesca Toni
Imperial College London, GB
- Leon van der Torre
University of Luxembourg, LU
- Serena Villata
Laboratoire I3S – Sophia
Antipolis, FR
- Vern R. Walker
Hofstra Univ. – Hempstead, US
- Zhe Yu
University of Luxembourg, LU



Managing Technical Debt in Software Engineering

Edited by

Paris Avgeriou¹, Philippe Kruchten², Ipek Ozkaya³, and
Carolyn Seaman⁴

- 1 University of Groningen, Groningen, NL, paris@cs.rug.nl
- 2 University of British Columbia, Vancouver, BC, CA, pbk@ece.ubc.ca
- 3 Carnegie Mellon University, Pittsburgh, PA, US, ozkaya@sei.cmu.edu
- 4 University of Maryland, Baltimore County, MD, US, cseaman@umbc.edu

Abstract

This report documents the program and outcomes of Dagstuhl Seminar 16162, “Managing Technical Debt in Software Engineering.” We summarize the goals and format of the seminar, results from the breakout groups, a definition for technical debt, a draft conceptual model, and a research road map that culminated from the discussions during the seminar. The report also includes the abstracts of the talks presented at the seminar and summaries of open discussions.

Seminar April 17–22, 2016 – <http://www.dagstuhl.de/16162>

1998 ACM Subject Classification coding tools and techniques, design tools and techniques, management, metrics, software engineering

Keywords and phrases software decay, software economics, software evolution, software project management, software quality, technical debt

Digital Object Identifier 10.4230/DagRep.6.4.110

Edited in cooperation with Robert Nord

1 Executive Summary

Ipek Ozkaya

Philippe Kruchten

Robert Nord

Paris Avgeriou

Carolyn Seaman

License © Creative Commons BY 3.0 Unported license
© Ipek Ozkaya, Philippe Kruchten, Robert Nord, Paris Avgeriou,
and Carolyn Seaman

The term *technical debt* refers to delayed tasks and immature artifacts that constitute a “debt” because they incur extra costs in the future in the form of increased cost of change during evolution and maintenance. The technical debt metaphor provides an effective mechanism for communicating design trade-offs between developers and other decision makers. When managed effectively, technical debt provides a way to gauge the current maintainability of a system and correct the course when that level is undesirable. While other software engineering disciplines – such as software sustainability, maintenance and evolution, refactoring, software quality, and empirical software engineering – have produced results relevant to managing technical debt, none of them alone suffice to model, manage, and communicate the different facets of the design trade-off problems involved in managing technical debt.



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

Managing Technical Debt in Software Engineering, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 110–138

Editors: Paris Avgeriou, Philippe Kruchten, Ipek Ozkaya, and Carolyn Seaman



DAGSTUHL
REPORTS

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

Despite recent progress by the research community in understanding technical debt, increased attention by tool vendors on assessing technical debt through code conformance checking, and collaboration with industry in sharing data and challenges, there are several open questions about the role of technical debt in software development. The goal of this seminar was to establish a common understanding of key concepts of technical debt and build a road map for future work in this area to address these open questions.

How do we define and model technical debt? The software engineering community is converging on defining *technical debt* as making technical compromises that are expedient in the short term, but that create a technical context that increases complexity and cost in the long term. While the conceptual roots of technical debt imply an idealized, deliberate decision-making process and rework strategy as needed, we now understand that technical debt is often incurred unintentionally and catches software developers by surprise. Hence, it is mostly observed during maintenance and evolution. Technical debt as a metaphor serves as a strong communication mechanism, but the community now understands that technical debt is also a software development artifact. This overloaded nature creates confusion, especially for newcomers to the field. In addition, there is a risk of associating anything detrimental to software systems and development processes with technical debt. This risk necessitates crisply defining both technical debt and related concepts.

How do we manage technical debt? Managing technical debt includes recognizing, analyzing, monitoring, and measuring it. Today many organizations do not have established practices to manage technical debt, and project managers and developers alike are longing for methods and tools to help them strategically plan, track, and pay down technical debt. A number of studies have examined the relationship between software code quality and technical debt. This work has applied detection of “code smells” (low internal code quality), coupling and cohesion, and dependency analysis to identify technical debt. However, empirical examples collected from industry all point out that the most significant technical debt is caused by design trade-offs, which are not detectable by measuring code quality. Effective tooling to assist with assessing technical debt remains a challenge for both research and industry.

How do we establish an empirical basis and data science for technical debt? Well-defined benchmarks provide a basis for evaluating new approaches and ideas. They are also an essential first step toward creating an empirical basis on which work in this area can grow more effectively. Effective and well-accepted benchmarks allow researchers to validate their work and tailor empirical studies to be synergistic. Technical debt’s evolving definition and its sensitivity to context have inhibited the development of benchmarks so far. An ideal benchmark for technical debt research would consist of a code base, architectural models (perhaps with several versions), and known technical-debt items (TD items). New approaches to identify technical debt could be run against these artifacts to see how well the approaches reveal TD items. Industry needs guidance for how and what data to collect and what artifacts they can make available to enable progress in understanding, measuring, and managing technical debt.

Seminar Format

In this seminar, we brought together researchers, practitioners, and tool vendors from academia and industry who are interested in the theoretical foundations of technical debt and how to manage it from measurement and analysis to prevention. Before the seminar,

the organizers created a blog where attendees could post positions and start discussions to facilitate seeding of ideas.

Before the seminar, the organizers grouped discussions and blog entries into relevant themes that included creating a common definition and conceptual model of technical debt, measurement and analysis of technical debt, management of technical debt, and a research road map for managing technical debt.

Our goal was to make this seminar a working week; hence we had a dynamic schedule. We did not feature any long talks. Each day had three types of sessions. There was a plenary session for “lightning talks,” in which each presenter had 10 minutes for presentation and questions on each day except for the last day of the seminar. The second type of session was for breakout discussions. Breakout sessions focused on themes that emerged from the blog and the goals of the seminar. Participants first discussed these in randomly assigned small groups in order to maximize cross-pollination of ideas. Last, we had plenary discussion sessions to collate and summarize the discussions during the breakouts. At the end of each day, the organizers asked for feedback and adjusted the flow of the following day accordingly. As a result, we dedicated the fourth day of the seminar to an “un-conference” format in which the discussion topic emerged based on the interests and votes of the attendees. The summaries of these sessions are included in Section 5: Open Problems.

The Definition of Technical Debt and a Conceptual Model

At the conclusion of the seminar, attendees agreed on the following working definition of technical debt, which we refer to as the 16162 definition of technical debt:

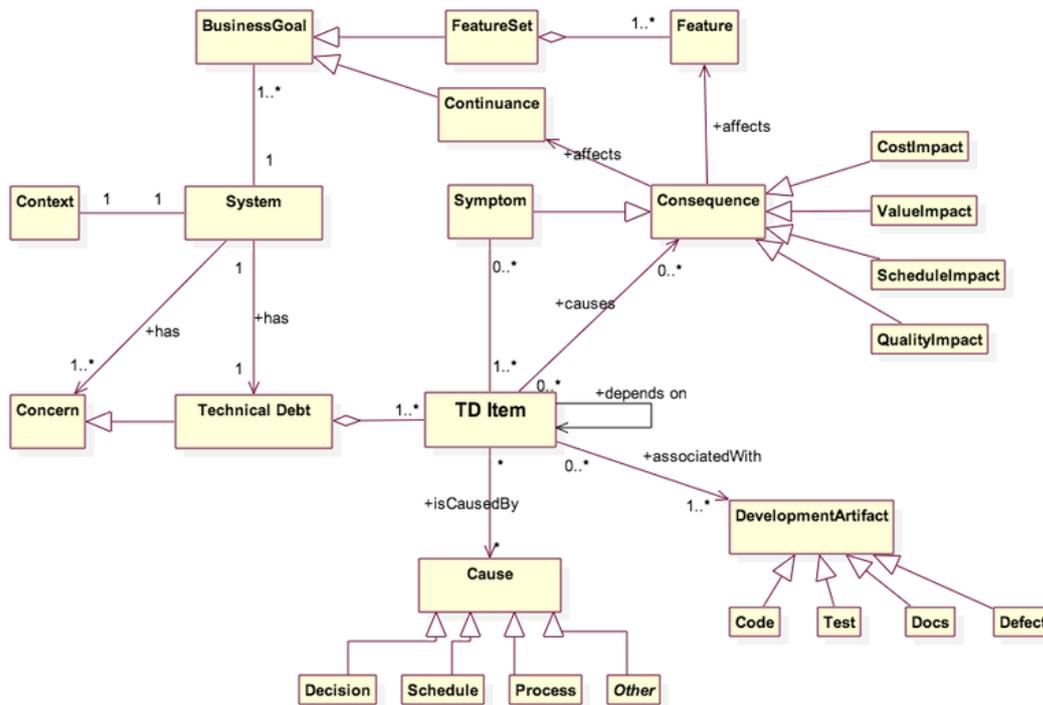
In software-intensive systems, technical debt is a collection of design or implementation constructs that are expedient in the short term, but set up a technical context that can make future changes more costly or impossible. Technical debt presents an actual or contingent liability whose impact is limited to internal system qualities, primarily maintainability and evolvability.

A significant outcome of the week was the recognition that, similar to other complex software engineering artifacts, technical debt is best described through multiple viewpoints. Concepts related to technical debt in particular should be discussed based on two related viewpoints:

1. the viewpoint describing the properties, artifacts, and elements related to technical debt items
2. the viewpoint articulating the management- and process-related activities to perform, or the different states that debt may go through

Figure 1 shows the initial conceptual model that served as the starting point for discussions. This model helped the group converge on key concepts. Mismatches occurred when the discussions focused on causes that may or may not be input to measurement and analysis. The dynamic view is intended to articulate these aspects.

The technical debt associated with a software-intensive system is composed of a set of TD items, and this technical debt is one of many concerns associated with a system. TD items have both causes and consequences. The cause of technical debt can be a process, a decision, an action (or lack thereof), or an event that triggers the existence of that TD item, such as schedule pressure, unavailability of a key person, or lack of information about a technical feature.



■ **Figure 1** Conceptual Model for Technical Debt.

The consequences of a TD item are many: technical debt can effect the value of the system, the costs of future changes, the schedule, and system quality. The business objectives of the sponsoring organization developing or maintaining the software system are affected in several ways: through delays, loss of quality for some features of the system, and difficulties in maintaining the system operations (continuance).

A TD item is associated with one or more concrete, tangible artifacts of the software development process, primarily the code, but also to some extent the documentation, known defects, and tests associated with the system.

To keep with the financial metaphor, the cost impact of technical debt can be seen as composed of principal and interest. The principal is the cost savings gained by taking some initial approach or shortcut in development (the initial principal, often the initial benefit) or the cost that it would now take to develop a different or better solution (the current principal).

The interest is comprised of costs that add up as time passes. There is recurring interest: additional cost incurred by the project in the presence of technical debt, due to reduced velocity (or productivity), induced defects, and loss of quality (maintainability is affected). And there are accruing interests: the additional cost of the developing new software depending on not-quite-right code (evolvability is affected).

This view summarizing the elements related to technical debt, however, does not capture causes that may or may not be input to measurement and analysis, the activities that need to be conducted to manage technical debt, and the states debt may go through. Another view is intended to articulate these aspects.

This definition and the model serve as the starting point for the community to build on and improve.

Research Road Map

One outcome of the seminar was a broad agenda for future work in technical debt research. While this road map needs to be fleshed out in the future with more detailed research questions and problem statements, it lays out three areas that require attention. First is the identification of a core concept – value – that is central to the technical debt metaphor and that needs definition and operationalization. Second is a recognition that there is an important context to technical debt that should be studied. There are attributes of the context of any particular instance of technical debt in a real environment that must be understood. But there are also other phenomena that are related to technical debt that should be studied, such as other types of “debt.” Third, the road map lays out the community’s basic infrastructure needs, which will enable further collaboration and progress in this area. The research road map that arose out of the discussions at Dagstuhl is described in more detail in this report under Section 4: Working Groups.

Follow-up Work

At the seminar, participants recognized that a carefully considered conceptual model and research road map would be useful outputs for the broader community interested in managing technical debt. Hence, more comprehensive explanation of a conceptual model and the research road map are planned as publications in appropriate venues once the community has a chance to vet the ideas further. The blog established before the seminar will continue to facilitate this interaction.

2 Table of Contents

Executive Summary

<i>Ipek Ozkaya, Philippe Kruchten, Robert Nord, Paris Avgeriou, and Carolyn Seaman</i>	110
--	-----

Overview of Talks

Technical Debt: Financial Aspects <i>Areti Ampatzoglou</i>	117
Towards a New Technical Debt Index: A Code and Architecture-Driven Index <i>Francesca Arcelli Fontana and Riccardo Roveda Marco Zanoni</i>	117
Dynamic and Adaptive Management of Technical Debt: Managing Technical Debt @Runtime <i>Rami Bahsoon</i>	118
Towards Measuring the Defect Debt and Building a Recommender System for Their Prioritization <i>Ayse Basar Bener</i>	119
Technical Debt Management in Practice <i>Frank Buschmann</i>	119
Relative Estimates of Technical Debt <i>Alexandros Chatzigeorgiou</i>	120
Measuring and Communicating the Technical Debt Metaphor in Industry <i>Bill Curtis</i>	120
On the Interplay of Technical Debt and Legacy <i>Johannes Holvitie and Ville Leppänen</i>	122
Technical Debt Aware Modeling <i>Clemente Izurieta</i>	122
Business Value of Technical Debt <i>Heiko Koziolok and Klaus Schmid</i>	123
Prioritization of Technical Debt <i>Antonio Martini and Jan Bosch</i>	123
Technical Debt in Scientific Research Software <i>John D. McGregor</i>	124
Google Experience Report Engineering Tradeoffs and Technical Debt <i>J. David Morgenthaler</i>	124
Technical Debt in Product Lines <i>Klaus Schmid</i>	125
On Concept Maps for TD Research <i>Carolyn Seaman</i>	126
Technical Debt Concepts in Architectural Assessment <i>Andriy Shapochka</i>	126
An Approach to Technical Debt and Challenges in the Acquisition Context <i>Forrest Shull</i>	127

From Technical to Social Debt and Back Again <i>Damian Andrew Tamburri and Philippe Kruchten</i>	128
Technical Debt Awareness <i>Graziela Tonin, Alfredo Goldman, and Carolyn Seaman</i>	128
Working Groups	
A Research Road Map for Technical Debt <i>Carolyn Seaman</i>	129
Open Problems	
The Interplay Between Architectural (or Model Driven) Technical Debt vs. Code (or Implementation) Technical Debt <i>Clemente Izurieta</i>	132
From Technical Debt to Principal and Interest <i>Andreas Jedlitschka, Liliana Guzmán, and Adam Trendowicz</i>	133
Community Datasets and Benchmarks <i>Heiko Koziolok and Mehdi Mirakhorli</i>	133
Deprecation <i>J. David Morgenthaler</i>	134
Report of the Open Space Group on Technical Debt in Product Lines <i>Klaus Schmid, Andreas Jedlitschka, John D. McGregor, and Carolyn Seaman</i> . . .	134
Automating Technical Debt Removal <i>Will Snipes and Andreas Jedlitschka</i>	135
Social Debt in Software Engineering: Towards a Crisper Definition <i>Damian Andrew Tamburri, Bill Curtis, Steven D. Fraser, Alfredo Goldman, Johannes Holvitie, Fabio Queda Bueno da Silva, and Will Snipes</i>	136
An Advanced Perspective for Engineering and Evolving Contemporary Software <i>Guilherme Horta Travassos</i>	136
Participants	138

3 Overview of Talks

3.1 Technical Debt: Financial Aspects

Areti Ampatzoglou (University of Groningen, NL)

License © Creative Commons BY 3.0 Unported license
© Areti Ampatzoglou

Joint work of Areti Ampatzoglou, Apostolos Ampatzoglou, Alexander Chatzigeorgiou, Paris Avgeriou
Main reference A. Ampatzoglou, A. Ampatzoglou, A. Chatzigeorgiou, P. Avgeriou, “The financial aspect of managing technical debt: A systematic literature review”, *Information & Software Technology*, Vol. 64, pp. 52–73, 2015.

URL <http://dx.doi.org/10.1016/j.infsof.2015.04.001>

The concept of technical debt is closely related to the financial domain, not only due to the metaphor that bonds it with financial debt, but also because technical debt represents money. On the one hand, it represents money saved while developing at a lower quality or money earned when delivering the product in time, whereas on the other hand, it represents money spent when applying a refactoring. As a result, financial terms are broadly used in TD literature. In order to work towards a framework for managing technical debt, we have attempted to organize a glossary of the most common financial terms that are used in the state of the art. The glossary presents these terms and a definition for each one. The definitions are a result of synthesizing the way that the terms are used in literature and in some cases they reflect our understanding of how these notions could prove beneficial for technical debt management. Additionally, we have illustrate our view on how financial terms are used in technical debt literature and the way they are linked to each other.

3.2 Towards a New Technical Debt Index: A Code and Architecture-Driven Index

Francesca Arcelli Fontana (University of Milano-Bicocca, IT) and Riccardo Roveda Marco Zanoni

License © Creative Commons BY 3.0 Unported license
© Francesca Arcelli Fontana and Riccardo Roveda Marco Zanoni

Main reference F. Arcelli Fontana, V. Ferme, M. Zanoni, R. Roveda, “Towards a Prioritization of Code Debt: A Code Smell Intensity Index”, in *Proc. of 7th IEEE International Workshop on Managing Technical Debt (MTD@ICSME’15)*, pp. 16–24, IEEE CS, 2015.

URL <http://dx.doi.org/10.1109/MTD.2015.7332620>

URL <http://essere.disco.unimib.it/>

In our laboratory of Evolution of Software SystEms and Reverse Engineering (ESSeRE Lab) of University of Milano Bicocca, we have experimented five different tools able to provide a Technical Debt Index, sometime called in different ways, but with the same or similar purpose. We found that, often, architectural issues are not taken into account, and when they are considered the main focus is on the detection of cyclic dependencies or other dependency issues. Many other architectural problems/smells are not considered, e.g., the relations (structural or statistical) existing among code and architectural problems. Moreover, different architectural smells/problems can be identified only by analyzing the development history of a system, and TD indexes do not take into account this kind of information too. We would like to work on the definition of a new TD index, with a focus on code and architectural debt, and experiment it on a large dataset of projects. In the TD index computation we would like to consider:

- Code and architectural smells detection
- Code and architecture/design metrics

- History of a system, including code changes, lifespan of the code and architectural smells
- Identification of problems more critical than others, to weight the collected analysis elements (e.g., metrics, smells, issues) according to their relevance in existing (past) projects

To this aim, we are working with colleagues of other two Universities on the definition of a catalogue of architectural smells (AS) and their classification. The classification could be useful to better explore possible relations existing among the architectural smells. We are currently working on the identification of some architectural smells by exploiting different metrics, the history of a system and their possible correlations (structural and/or statistical). The choice of thresholds considered for the applied metrics is a critical problem and can be determined through statistical analysis and/or regression (machine learning). Moreover, we have to identify the most critical problems/AS in order to prioritize their removal. The information on the code parts of a system that have been subjected to more changes in their history and, that we expect that will be more changed or extended in the future, can be used for the prioritization of the problems to be removed, together with the information on the code parts with low quality in terms of metric values and code/architectural smells.

3.3 Dynamic and Adaptive Management of Technical Debt: Managing Technical Debt @Runtime

Rami Bahsoon (University of Birmingham, GB)

License  Creative Commons BY 3.0 Unported license
© Rami Bahsoon

The talk has highlighted ongoing effort on managing technical debt in open, dynamic and adaptive environments, where we have looked at dynamic and adaptive composition of cloud-based architecture as an example. We have motivated the need for treating technical debt as a “moving target” that needs to be dynamically and adaptively monitored for prevention and/or transforming the debt into value. We have argued that the much of the debts can be linked to ill- and poorly- justified runtime decisions that can carry short-term gains but not geared for long-term benefits and future value creation. The debt can be observed on utilities linked to Quality of Services (QoS), Service Level violations, need for excessive and costly adaptation etc. The talk has highlighted examples of these decisions and has looked at two interconnected angles for managing debt at runtime: (i) predicative and preventative design support for debt-aware dynamic and adaptive systems and (ii) using online and adaptive learning as mechanisms for proactive runtime management of debts. We have revisited the conceptual technical debt model of Kruchten et al. to make runtime debt concerns, items, artefacts, consequences, etc. explicit. We have seen a need for enriching the model with time-relevant information to address pragmatic needs for runtime management of debt.

3.4 Towards Measuring the Defect Debt and Building a Recommender System for Their Prioritization

Ayse Basar Bener (Ryerson University – Toronto, CA)

License © Creative Commons BY 3.0 Unported license

© Ayse Basar Bener

Joint work of Shirin Akbarinasaji

Main reference S. Akbarinasaji, A. B. Bener, A. Erdem, “Measuring the Principal of Defect Debt”, in Proc. of the 5th Int’l Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE@ICSE’16), pp. 1–7, ACM, 2016.

URL <http://dx.doi.org/10.1145/2896995.2896999>

URL <http://www.ryerson.ca/~abener/pubs.html>

Due to tight scheduling and limited budget, the software development may not be able to resolve all the existing bugs in a current release. Similar to the concept of technical debt, there are also defects that may be postponed to be fixed in the upcoming releases. Such lingering defects are left in the system intentionally or unintentionally, but they themselves create debt in the system similar to the technical debt metaphor. In this research, we particularly focus on defect debt. The accumulation of the deferred bugs in the issue tracking system leads to the rise of the defect debt. In order to manage the defect debt, software managers require to be aware of amount of debt (principal) and interest in their system. There are several studies in the literature which measure the principal, however, only few researchers investigate the interest amount. In this study, we propose two novel approaches to calculate the interest amount for defect debt based on the severity, priority of defect and graph theory analysis. We then propose a dynamic model such as reinforcement learning that learns dynamically from its environment in order build a recommender system to prioritize defects based on the debt (principal and interest).

3.5 Technical Debt Management in Practice

Frank Buschmann (Siemens AG – München, DE)

License © Creative Commons BY 3.0 Unported license

© Frank Buschmann

In real-world development Technical Debt is always present. Technical Debt Management is thus a continuous activity; almost on a day-to-day basis. This puts the following four requirements on any Technical Debt management environment: Technical Debt identification and assessment must be performed automatically. Appropriate tools must retrieve and handle information from multiple sources, such as code, backlogs, and architecture documentation. A defined quality model is necessary to assess whether or not Technical Debt has actually occurred or is beyond a defined threshold. Identified Technical Debt must be put automatically into the context of development to decide where in the system it is of value to manage it. For instance, Technical Debt in code to be extended for the next product release is more critical than Technical Debt in system areas that stay untouched. Technical Debt to manage must be prioritized according to concrete release goals. For instance, Technical Debt that causes rippling effects to other modules it is likely of higher priority than Technical Debt that stays internal to a module. Measures to address Technical Debt must balance effort and value. Features are a system’s asset, code is its liability. In the end it is important to deliver a competitive system, not a debt-free system. It is also important to minimize the root causes of Technical Debt – regardless of whether it is taken consciously or accidentally Development

processes should be freed from practices that put technical challenges on development teams or invites them to game with the process. Architecture styles like Microservices reduce the occurrence and even more the outreach of Technical Debt in a system. Technical trainings for teams on deliberate design and coding practices can substantially limit the occurrence of Technical Debt.

3.6 Relative Estimates of Technical Debt

Alexandros Chatzigeorgiou (University of Macedonia – Thessaloniki, GR)

License © Creative Commons BY 3.0 Unported license

© Alexandros Chatzigeorgiou

Joint work of Ampatzoglou, Areti; Ampatzoglou, Apostolos; Avgeriou, Paris; Amanatidis, Theodoros
Main reference A. Ampatzoglou, A. Ampatzoglou, A. Chatzigeorgiou, P. Avgeriou, “The Financial Aspect of Managing Technical Debt: A Systematic Literature Review”, *Information and Software Technology*, Vol. 64, pp. 52–73, 2015.

URL <http://dx.doi.org/10.1016/j.infsof.2015.04.001>

Industry and academia agree that technical debt has to be managed and assessed. However, the TD community lacks commonly agreed methods/tools for TD estimation. Almost all current approaches for TD measurement are based on the identification of individual inefficiencies, usually at the design and code level. However, developers are reluctant to accept as a panacea any approach that sets arbitrary thresholds for the desired quality. To address this practical challenge we argue that design/code inefficiencies should not be assessed against ‘hard’ thresholds but on the basis of relative measurements. We believe that the theory and practice of search-based software engineering (SBSE) can be exploited to extract a design that optimizes a selected fitness function. The ‘distance’ between an actual design and the corresponding optimum one can serve as an estimate of the effort that has to be spent to repay TD. Although the notion of an ‘optimum’ system might sound utopic, the benefit from the use of a relative estimate is twofold: a) it expresses a potentially achievable level of TD, b) the aforementioned distance can be mapped to actual refactoring activities. A side benefit of such a quantification approach is that it allows the assessment of individual developer contribution to TD. Through the analysis of software repositories we can assess the impact of individual commits. This information could be exploited to compile individual TD reports for open-source contributors and facilitate the collection of information by providing an additional motivation for participating in surveys related to TD.

3.7 Measuring and Communicating the Technical Debt Metaphor in Industry

Bill Curtis (CAST – Fort Worth, US)

License © Creative Commons BY 3.0 Unported license

© Bill Curtis

Joint work of Douziech, Philippe-Emmanuel; Curtis, Bill

Managers and executives across industry have embraced the technical debt metaphor because it describes software issues in a language that industry understands. However, their tacit understanding of the metaphor differs from the typical formulation in the technical debt research community. Whereas researchers often limit this metaphor to suboptimal design choices that should eventually be corrected, most in industry think of technical debt as the

collection of software flaws that need to be corrected regardless of their type. Industry wants the principal of a technical debt to estimate their corrective maintenance expense and total cost of ownership. Thus, industry is using the technical debt metaphor to describe phenomena it needs to quantify in financial terms. Industry does not care whether a structural flaw fits Cunningham's original concept of technical debt. Rather it cares that IT must spend money to correct these flaws, and that there may be a marginal cost in inefficient use of human and machine resources (i.e., interest) until they are fixed.

Going forward there will probably be at least two divergent perspectives on technical debt, each proper and valuable for its intended use. The research community will focus on sub-optimal design decisions that primarily affect the maintenance and evolvability of software. The industrial community will take a broader perspective on the flaws categorized as technical debt in order to use the metaphor to communicate the cost of software quality problems to the business in terms the business understands. These divergent views can co-exist because they serve different purposes and audiences. If technical debt were limited to architectural smells, industry would eventually turn to other concepts for explaining its broader issues in financial terms.

Measuring technical debt as an estimator of corrective maintenance costs implies measuring software flaws that must be fixed. If a flaw is not sufficiently damaging that its correction can be deferred endlessly, it is not an item of technical debt because the organization does not plan to spend money correcting it. The Consortium for IT Software Quality (CISQ) has defined measures of structural quality related to Reliability, Security, Performance Efficiency, and Maintainability based on counting structural flaws at the architectural and code unit level that can be detected through static analysis. The criterion for including a structural flaw in calculating these measures was that a panel of industry experts had to consider it severe enough to require correction. For instance, the CISQ Security measure was constructed from the Top 25 CWEs (Common Weakness Enumeration, the basis for the SANS Top 25 and OWASP Top 10) that hackers exploit to gain unauthorized entrance into a system.

The CISQ measures not only provide estimates of corrective maintenance costs, they also provide indicators related to the risk that systems can experience outages, data corruption, performance degradation, and other operational problems. The CISQ measure of technical debt will aggregate the structural flaws in the four quality characteristic measures and apply an estimate of corrective effort for each. When aggregated across the four CISQ measures, this measure should provide a good estimator of corrective maintenance costs as well as operational risks. The measure can be modified as empirical results from industry indicate improvement opportunities. Industry's willingness to participate in empirical validation studies is crucial for validating and improving the explanatory power of the technical debt metaphor.

3.8 On the Interplay of Technical Debt and Legacy

Johannes Holvitie (University of Turku, FI) and Ville Leppänen

License © Creative Commons BY 3.0 Unported license
© Johannes Holvitie and Ville Leppänen

Joint work of Holvitie, Johannes; Hyrynsalmi, Sami; Leppänen, Ville

Main reference J. Holvitie, V. Leppänen, S. Hyrynsalmi, “Technical Debt and the Effect of Agile Software Development Practices on It – An Industry Practitioner Survey”, in Proc. of the 6th Int’l Workshop on Managing Technical Debt (MTD@ICSME’14), pp.35–42, IEEE CS, 2014.

URL <http://dx.doi.org/10.1109/MTD.2014.8>

Different methods for technical debt accumulation have been discussed, but they have mainly focused on immediate accumulation. Arguably, there is also delayed accumulation, wherein the environment around static software assets changes. Here, updates to the assets no longer deliver the environments assumptions to the assets, they become detached from the environment, and we find the assets to have accumulated debt. This debt closely reminds of software legacy, as it represents assets which can no longer be subjected under the same development actions as newly created ones. In a recent multi-national survey, it was discovered that over 75% of technical debt instances’ have perceived origins in software legacy. This communicates about the close relation between the legacy and the debt concepts. While it encourages us to further explore applying established legacy management methods for technical debt, the relation also exposes challenges. Notably, we should consider if legacy is being merely rebranded into the more favorable technical debt. And if this is the case, how do we ensure that all aspects of the legacy instances are identified so as to convert them into fully manageable technical debt assets. Failure to do so, will result into having technical debt assets with varying levels of accuracy, and this undoubtedly hinders technical debt management efforts overall. Nevertheless, while considering these challenges, the software legacy domain should be further researched as both a commonality of technical debt instances and as a possible interface for enhancing existing management approaches.

3.9 Technical Debt Aware Modeling

Clemente Izurieta (Montana State University – Bozeman, US)

License © Creative Commons BY 3.0 Unported license
© Clemente Izurieta

Joint work of Rojas, Gonzalo

Main reference I. Griffith, D. Reimanis, C. Izurieta, Z. Codabux, A. Deo, and B. Williams, “The Correspondence Between Software Quality Models and Technical Debt Estimation Approaches”, in Proc. of the 6th International Workshop on Managing Technical Debt (MTD@ICSME’14), pp. 19–26, IEEE CS, 2014.

URL <http://dx.doi.org/10.1109/MTD.2014.13>

The Software Engineering Laboratories (SEL) at Montana State University has been engaged in active Technical Debt research for approximately five years. Our research and development goal is to find a balance between practical applications of research findings in the form of very useful tools tailored for commercial customers. We are currently developing dashboard technology that builds on the SonarQube framework infrastructure to provide functionality that calculates the quality of software according to various ISO standards, as well as providing technical debt measurements. The framework is tailored to be extensible by allowing for additional plug-ins. For example, we are currently focusing on building the Risk Management Framework quality model (RMF). SEL is also focusing on researching the various modeling decisions that affect the technical debt of associated code generated from such models.

This research will lead to a plug in that is tailored for modelers and architects, but that is technical debt aware by providing different model smell refactoring alternatives that a modeler may choose from. Each choice has decisively different outcomes in the technical debt measurements performed in the corresponding generated code. This line of research will lead to better traceability and to the understanding of the relationship between architecture and code.

3.10 Business Value of Technical Debt

Heiko Kozirolek (ABB AG Forschungszentrum – Ladenburg, DE) and Klaus Schmid (Universität Hildesheim, DE)

License © Creative Commons BY 3.0 Unported license
© Heiko Kozirolek and Klaus Schmid

Technical debt prioritization is challenging, because the benefit of resolving technical debt (TD) items varies depending on the planned evolution of a software system. Fixing TD items in code parts that will not be modified provides no immediate benefits. The decision on how many TD items to resolve depends on the business context a software system is developed in. In an innovative, rapidly growing market, time-to-market delivery may be essential for product success, so taking on technical debt in these situations may be warranted. These observations call for more emphasis on modeling the business context of software systems, capturing the view of product managers and decision makers. However, most TD analysis and resolution methods and tools are developed from the perspective of the software engineers and architects, not explicitly accounting for market analysis or future evolution scenarios. Therefore it appears useful to develop method and tools in collaboration with both developer and decision makers. We envision an Integrated Technical Debt Analysis Environment that can integrate both the output of software artifact analysis tools as well as the output of modelling the business context and development road maps. Such an instrument would allow to make an informed decision about taking or resolving technical debt respecting both the developer and management perspective.

3.11 Prioritization of Technical Debt

Antonio Martini (Chalmers UT – Göteborg, SE) and Jan Bosch

License © Creative Commons BY 3.0 Unported license
© Antonio Martini and Jan Bosch

Joint work of Antonio Martini, Jan Bosch

Main reference A. Martini, J. Bosch, “An Empirically Developed Method to Aid Decisions on Architectural Technical Debt Refactoring”, in Proc. of the 38th Int’l Conf. on Software Engineering (ICSE’16) – Companion Volume, pp. 31–40, ACM, 2016.

URL <http://dx.doi.org/10.1145/2889160.2889224>

Main reference A. Martini and J. Bosch, “Towards Prioritizing Architecture Technical Debt: Information Needs of Architects and Product Owners”, in Proc. of the 41st Euromicro Conf. on Software Engineering and Advanced Applications (EUROMICRO-SEAA’15), pp. 422–429, IEEE CS, 2015.

URL <http://dx.doi.org/10.1109/SEAA.2015.78>

A Technical Debt item needs to be prioritized against features and among other TD items. There is a need for mechanisms, methods and tools to aid the stakeholders in prioritizing the refactoring (repayment) of Technical Debt. An important step is to understand the Technical Debt impact (interest). We found that the estimated impact of TD items provides useful

information when the stakeholders (technical and non-technical) prioritize with respect to aspects such as Lead Time, Maintenance Cost and Risk. However, there are aspects that are considered of higher priority in commercial software organizations, such as Competitive Advantage, Specific Customer Value and Market Attractiveness. It is important to understand if and how the impact of Technical Debt affects such aspects (directly or indirectly). In order to understand if and when a Technical Debt item should be refactored with respect to other items, we proposed an approach based on the calculation of the ratio Principal/Interest. Such an approach (AnaConDebt) is based on a checklist of key factors that compose Principal and Interest. The assignment of weights to such factors, based either on expert experience or on metrics available at the organizations, gives a result that is simple to interpret and useful for comparison. The comparison of the ratio with other TD items' ratios helps the prioritization. Multiple ratios can also be calculated at different points in time for the same item, to estimate if the refactoring can be postponed or not. This approach can be repeated iteratively to adjust and reprioritize the refactoring according to new information available to the stakeholders.

3.12 Technical Debt in Scientific Research Software

John D. McGregor (Clemson University, US)

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

Developing scientific research software is expanding rapidly as research methods are more software-based. The process for developing that software is a breeding ground for technical debt. In most research groups the students of physics, chemistry or other discipline are not trained in software development and the managers of those students are professors with even less software experience than their younger students. The problems with these systems can be considered technical debt because the work of that research group depends on building on the software built by previous students. Research progress is slowed if that existing software must be modified before new work can be started. One research study recently was delayed due to memory limitations. They ported to a larger system only to find there were inherent assumptions throughout the software that limited the ability to take advantage of the larger system. Our work on scientific software ecosystems is investigating how to establish and maintain a supply chain of software which allows students to select software that will be suitable for their work.

3.13 Google Experience Report Engineering Tradeoffs and Technical Debt

J. David Morgenthaler (Google Inc. – Mountain View, US)

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

Software development at Google sits at one extreme on the spectrum of development environments, and as such may not represent the challenges typically seen by the vast majority of other companies, developers, or projects. Yet I believe the issues Google faces with rapid technical evolution of our heavily reused internal components point to the future

direction the industry as a whole is heading. Google's underlying infrastructure, whether hardware or software, is changing so quickly that developers often face a new type of tradeoff – between platforms, frameworks and libraries that are stable but superseded, and those that are new and state-of-the-art, but currently incomplete. The up-front design decision is not whether to take on technical debt, but which form. Either speedy development using tried and true, but soon to be unsupported, platforms, or slower, more painful, work as a guinea pig for the next cool development approach with its promise of a longer life span.

Google also imposes these debt decisions on external developers who use open-sourced Google platforms. The Android OS is one well-documented example, with major revisions shipping nearly every year. This fast pace of innovation also leads to version fragmentation and rapid software aging, and therefore higher maintenance costs. The road ahead is indeed fraught with danger. Yet with an 'installed base' of 1.5 billion active devices and 2 million available apps, this ecosystem represents a tremendous opportunity, both for developers to reach billions of future users, and for future technical debt research.

3.14 Technical Debt in Product Lines

Klaus Schmid (Universität Hildesheim, DE)

License © Creative Commons BY 3.0 Unported license
© Klaus Schmid

Joint work of Sascha El-Sharkawy, Adam Krafczyk

Main reference S. El-Sharkawy, A. Krafczyk, and K. Schmid, "Analysing the Kconfig Semantics and Its Analysis Tools", in Proc. of the 2015 ACM SIGPLAN Int'l Conf. on Generative Programming: Concepts and Experiences (GPCE'15), pp. 45–54, ACM, 2015.

URL <http://dx.doi.org/10.1145/2814204.2814222>

Traditionally technical debt (TD) research focused particular on individual systems and the research on managing TD aimed at supporting a single project. However, it should be recognized that also product lines may contain (significant) forms of TD. – And what is more, this TD may come in new and different forms that are not yet addressed by existing tools. In our research we are currently focusing on technical debt in software product lines. Ideally, product lines consist of a variability model along with corresponding assets that are managed to create the resulting product. In our research we focus in particular on logical anomalies that complicate the product line realization. Examples of this are dead code, undead code or over-constrained code that reduces the potential number of code configurations below the range of configurations described by the variability model. This is of course not the only way to address the problem of smells – similar to metrics in traditional technical debt research one could also search for code smells using metrics. A complexity that arises in particular in product line research is that technical debt related to configurations cannot be addressed without taking also into account the build space. Hence, we need to integrate information from variability model, code variability, and build information to arrive at reasonable conclusions regarding product technical debt.

3.15 On Concept Maps for TD Research

Carolyn Seaman (University of Maryland, Baltimore County, US)

License  Creative Commons BY 3.0 Unported license
© Carolyn Seaman

From my point of view, the motivation for developing a concept map for the TD research community is two-fold:

- To standardize terminology.
- To aid in categorization of new and existing research.

However, there are other relevant motivations for concept mapping. I presented in this talk three concept maps that I have found in the TD literature, that are very different and have different motivations, in an effort to convey the breadth of ideas we should consider when developing a concept map for our community.

The first concept map I presented was a very old, very simple framework that I have used to organize work in my own lab. It divides TD research into TD identification approaches, TD measurement and analysis approaches, approaches to making decisions using TD information and approaches to organizing and storing information about TD. While I have found the distinction between these areas of TD research, I think the community could use a more comprehensive and sophisticated model.

The second model I presented was derived from an in-depth historical case study of the events and decisions related to a single instance of TD. The goal was a grounded descriptive model of TD decision making. It's described in detail in a paper currently under review (Siebra, et al.).

The third model I talked about was based on an in-depth analysis of early (pre-2011), mostly non-scholarly, TD literature (Tom et al., 2013). The resulting map provides a useful categorization of TD precedents, TD outcomes, TD attributes, and software development dimensions (e.g. development phases).

3.16 Technical Debt Concepts in Architectural Assessment

Andriy Shapochka (SoftServe – Lviv, UA)

License  Creative Commons BY 3.0 Unported license
© Andriy Shapochka

An architectural assessment has become one of the most important tools in the architect's toolbox. Built on a mature methodologies such as ATAM and CBAM it serves well to evaluate software architectures and their implementations against the set of high priority quality attributes, constraints, and other architectural drivers relevant to the assessed system.

The assessment process is well defined, timeboxed to a few weeks and can be applied to the system in any phase of the software development lifecycle. It combines both qualitative and quantitative analysis techniques leading to a set of architecture improvement recommendations.

Architectural assessments are exceptionally well suited to evaluate maintainability and evolvability of the system at hand which essentially means its technical debt analysis including search for and categorization of potential technical debt items on the levels from overall system architecture to component design, to implementation details, their prioritization, value

for cost analysis of technical debt elimination, improvement metrics setup and monitoring, and other important activities.

As technical debt can manifest itself on different levels of abstraction and in other system perspectives such runtime quality attributes (performance, reliability, security, etc.) affected by the technical debt related tradeoffs, testability influenced by the complexity of the system implementation, metrics related to defects (frequencies, time to fix, defect density in components, etc.) proper analysis for the technical debt often involves comprehensive evaluation of the entire system architecture, its history, and plans for evolution.

There are multiple challenges still in need of addressing in the context of the technical debt assessments: quick and efficient semi-automated localization of the technical debt items, consistent value for cost analysis methodology meaningful and understood by business, technical debt interpretation and prioritization correlated with the business needs.

3.17 An Approach to Technical Debt and Challenges in the Acquisition Context

Forrest Shull (Carnegie Mellon University – Pittsburgh, US)

License © Creative Commons BY 3.0 Unported license
© Forrest Shull

This talk demonstrated a metrics-driven approach to TD management which has been used effectively with development organizations. The approach elicits quality goals and rules that the team feels will help achieve those goals, and for which compliance could be detected in the codebase. In an example case study, the quality goal was to minimize maintenance costs and the associated rule was to follow a reference architecture. Tools were used to search for rule violations (i.e. inconsistencies with the reference architecture) through all prior commits in the codebase, and to associate data from the change and defect tracker. Thus, the team could understand how often those rules were broken in the past and whether TD symptoms (e.g. increased change cost, decreased velocity) were correlated with noncompliance. If not, then the rules should be refined or replaced by new rules with more impact. One result from this work is the indication that no one set of TD detection rules consistently correlates with important TD symptoms across different projects and quality goals. Understanding TD in software that is being acquired / purchased presents unique challenges that are not amenable to such methods, however. In such instances the acquiring organization needs to understand the amount of extant technical debt, to estimate future sustainment costs: Will a substantial part of every dollar devoted to new functionality instead go to dealing with accumulated debt? Yet, for acquired software, many of the necessary data sources are not available for analysis. The talk ended by discussing the approaches relevant for acquisition, which need to include manual analysis. For further reading:

References

- 1 Nico Zazworka, Victor Basili, and Forrest Shull. *Tool Supported Detection and Judgment of Nonconformance in Process Execution*. 3rd International Symposium on Empirical Software Engineering and Measurement (ESEM), 2009
- 2 Jan Schumacher, Nico Zazworka, Forrest Shull, Carolyn B. Seaman, Michele A. Shaw. *Building Empirical Support for Automated Code Smell Detection*. ESEM 2010

- 3 Nico Zazworka, Michele A. Shaw, Forrest Shull, Carolyn Seaman. *Investigating the Impact of Design Debt on Software Quality*. MTD2011: 2nd Workshop on Managing Technical Debt,” ICSE, 2011, Honolulu, Hawaii, USA

3.18 From Technical to Social Debt and Back Again

Damian Andrew Tamburri (Polytechnic University of Milan, IT) and Philippe Kruchten (University of British Columbia – Vancouver, CA)

License © Creative Commons BY 3.0 Unported license
© Damian Andrew Tamburri and Philippe Kruchten

Main reference D. A. Tamburri, P. Kruchten, P. Lago, H. van Vliet, “Social debt in software engineering: insights from industry”, *Journal of Internet Services and Applications*, 17 pages, Springer, 2015; available open access.

URL <http://dx.doi.org/10.1186/s13174-015-0024-6>

An established body of knowledge discusses the importance of technical debt for product quality. In layman’s terms, Technical debt represents the current state of a software product as a result of accumulated technical decisions (e.g., architecture decisions, etc.). Techniques to study product quality limitations inherent to technical debt include technical data mining, code-analysis.

Nevertheless, quite recently we figured out that technical debt is only one face of the coin. In fact, social debt, the social and organisational counterpart to technical debt, plays a pivotal role in determining software product and process quality as much as technical debt helps phrase the additional technical project cost. Social debt reflects sub-optimal socio-technical decisions (e.g., adopting agile methods or outsourcing) which can compromise the quality of software development communities, eventually leading to software failure.

Studying and harnessing social debt is paramount to ensure successful software engineering in large-scale software development communities. I focus on explaining the definition and relations between technical and social debt while delineating techniques inherited from technical debt, organisational research and social networks research that could be rephrased to investigate social debt in software engineering (e.g., socio-technical code analysis, social-code graphs, socio-technical debt patterns, etc.).

3.19 Technical Debt Awareness

Graziela Tonin (University of Sao Paulo, BR), Alfredo Goldman (University of Sao Paulo, BR), and Carolyn Seaman (University of Maryland, Baltimore County, US)

License © Creative Commons BY 3.0 Unported license
© Graziela Tonin, Alfredo Goldman, and Carolyn Seaman

We conducted a study with teams of students (60 people total) on technical debt awareness. We asked the participants about the effects of explicitly identifying and tracking technical debt on their projects. Awareness of potential technical debt items sometimes led all team members to avoid contracting the debt. They often discussed such decisions, thus improving the team communication. As a result, several team members were more comfortable sharing their difficulties. Consequently, some teams had fewer ‘untouchable’ experts, and thus worked better as a real team. Decisions made to contract a debt were explicit and strategic, e.g. to deliver a new version to the customer faster. Moreover, making the technical debt list

visible allowed them to negotiate more time with the customer to perform refactoring. The technical debt list was used as a historical memory of the immature parts of the project by all teams. They could use it to check if there was debt to be paid in parts of the project that were about to be modified. The list was also a good indicator of the health of the project; it showed if code quality was improving or not. The teams created a culture of continuous improvement. In summary, making technical debt explicit had direct implications on the team's behavior. The majority of the team members said that, after they became aware of technical debt, they communicated more with other team members, thought more about the real need to contract debt, discussed in more deep about code quality, refactored more frequently, and understood the problems in the project better.

4 Working Groups

4.1 A Research Road Map for Technical Debt

Carolyn Seaman (University of Maryland, Baltimore County, US)

License © Creative Commons BY 3.0 Unported license
© Carolyn Seaman

Main reference N. Brown, Y. Cai, Y. Guo, R. Kazman, M. Kim, P. Kruchten, E. Lim, A. MacCormack, R. L. Nord, I. Ozkaya, R. S. Sangwan, C. B. Seaman, K. J. Sullivan, N. Zazworka, “Managing Technical Debt in Software-Reliant Systems”, in Proc. of the FSE/SDP workshop on Future of Software Engineering Research, pp. 47–52, ACM, 2010.

URL <http://dx.doi.org/10.1145/1882362.1882373>

Main reference P. Avgeriou, P. Kruchten, R. L. Nord, I. Ozkaya, C. B. Seaman, “Reducing Friction in Software Development”, IEEE Software, Vol. 33(1), pp. 66–73, 2016.

URL <http://dx.doi.org/10.1109/MS.2016.13>

At the end of a week of discussions about many aspects of technical debt (TD) research – past, ongoing, and envisioned – the attendees of the Dagstuhl workshop spent a morning sharing ideas about the most important TD-related research problems for the community to work on. The discussion of a research road map was grounded in the definition of TD formulated during the seminar:

In software-intensive systems, technical debt is a collection of design or implementation constructs that are expedient in the short term, but set up a technical context that can make future changes more costly or impossible. Technical debt presents an actual or contingent liability whose impact is limited to internal system qualities, primarily maintainability and evolvability.

The group spent some time envisioning how the world would be different if all our research efforts in this area were successful. This vision included the following points:

- TD would be managed as carefully as we currently manage defects and new features.
- We would have a clear, operational definition of “good-enough” software, including how much TD is acceptable.
- We would have a way to translate between developer concerns and manager concerns, and this translation would form a basis for making decisions about allocating effort to various tasks.
- TD would be incurred intentionally most of the time.
- Projects that manage TD would be more efficient, effective, and sustainable than projects that don't.

- The notion that up-front architectural work (vs. emergent architecture) is worth it would be well supported and accepted.
- There would be tools to support all aspects of TD management that are adopted and used by all stakeholders.
- TD-aware development (practices and tools) would be an accepted standard way of producing software.
- Architectural assessment would be part of standard development policy.

Achieving this vision will require changing development practices through effective communication between research and practice and, most importantly for the research community, showing the effectiveness of proposed approaches. The research community must also recognize that our practitioner audience consists of two different groups: managers who make decisions about spending money on eliminating TD and developers who write and maintain code. Proposed solutions must ultimately appeal to both, but researchers should also be clear about which group is primarily targeted with a new approach or tool.

The group discussed research activities in three broad areas, which are outlined in the sections below:

1. the core: defining, understanding, and operationalizing the concept of value with respect to TD
2. the essential context: understanding phenomena that fall outside the core definition of TD and that have an essential relationship with how TD plays out in practice
3. the necessary infrastructure: building the shared infrastructure that facilitates all our research activities

The Core: Understanding Value

The concept of value and what it means with respect to TD came up often and in many contexts throughout the seminar. It is not clear that we all mean the same thing when we talk about value, but it is clear that value is central to delivering effective mechanisms for managing TD in practice. Value comes into play both when deciding whether to incur TD (i.e., the value of a proposed TD item) as well as when deciding whether, when, and which TD to pay off (i.e., the value of eliminating an existing TD item). At one level, the concepts of principal and interest capture the short-term benefits (principal) and long-term costs (interest) of incurring TD, as well as the cost/benefit analysis related to paying it off. But these concepts do not adequately capture all aspects of value.

One aspect of value that is not easily captured by the concepts of principal and interest is the opportunity cost. One common benefit of incurring TD is the ability to take advantage of an opportunity (because resources are freed up) that might not be available at another time. In theory, this benefit is part of the principal of the TD being considered, but principal does not capture the time element (the fact that the opportunity is time-sensitive) nor does it allow for the fact that the resources freed by incurring TD and the resources needed to take advantage of the opportunity might not be comparable.

Another aspect of value that is even harder to make concrete is the value of TD management, and TD-related information, to the quality of decision making. Capturing the “quality” of decisions is difficult to begin with, but somehow demonstrating the benefits of considering TD in management decisions is a key area for TD researchers. A related question is how to define good-enough software. When does software have a level of quality (and a level of TD) that cannot be cost-effectively improved? Helping managers define “good enough” in their

context would be extremely useful, and better defining the value of TD is an important step in this process.

The quantification of value (and other TD properties) has always been an important aspect of TD research and was part of the research agenda set out in the first TD research road map in 2010 (Brown et al., 2010). Efforts to quantify principal and interest, and to somehow combine them into a measure of “value,” have proved difficult, but continued work in this area is crucial to furthering our vision of effective TD management in practice. Useful measures would both be at a level of granularity that is useful to developers in tracking and monitoring activities and be understandable in a business sense.

Steps toward providing effective operationalizations of value include proposing proxies for value, collecting data based on these proxies, designing validation procedures (i.e., study designs) to show that these measures are meaningful and useful, and conducting case studies to carry out these validations. Case studies can also be used to explore existing notions of value in practice, which would ground the choice of proposed metrics. Such exploratory case studies could also be used to discover grassroots approaches to TD management – that is, environments without an explicit strategy to manage TD, but where, out of necessity, techniques have emerged to deal with the most important aspects.

The Essential Context: Understanding Related Phenomena

When the Dagstuhl group agreed on our definition of TD, we also recognized that the definition does not encompass all topics that are important to study. TD exists in a context, and it exists in a variety of forms.

The TD definition limits TD to phenomena closely tied to source code (e.g., design debt, code debt, architecture debt), which leaves out other important types of debt that are at least partially analogous to financial debt, such as social debt, people debt, process debt, and infrastructure debt. These other types of debt are related to TD in various ways. Some are part of the causal chain leading to TD; the impacts of these other types of debt can lead to TD as defined. Others appear to co-occur with TD or create an environment that affects the ability to manage TD in some way.

Other aspects of TD context that need to be studied within the TD research agenda include

- uncertainty: building representations of uncertainty into TD value measures and using that information about uncertainty in decision making
- development and organizational context: studying what aspects of context affect how TD is most effectively managed
- time: better accounting for the passage of time (measured not necessarily in time units but in relevant events, such as code changes) in all aspects of TD management, such as the value of TD over time and the effect of time-based events on value
- dependencies and interactions: between TD items, between TD items and development artifacts
- knowledge management: what TD-related information to capture and disseminate for effective TD decision making
- causal chains: the constellation of “causes” leading up to the creation of TD
- TD maturity: the creation of a maturity model that depicts the different levels at which TD could be managed, or possibly incorporating TD concerns into existing maturity models

The Necessary Infrastructure: Data, Tools, and Other Components

One advantage of forming a coordinated, active research community in a particular area of study is the ability to share resources. The group at Dagstuhl identified a number of resources that members of the community could develop and contribute to the TD research infrastructure to enhance the level of progress of the entire field. Identifying relevant pieces of infrastructure is not possible until a community reaches a certain level of maturity. We feel that the TD research community has reached that level and that identifying and building common infrastructure would now be useful. This common infrastructure should include the following components:

- data sets: OSS projects provide some useful data, but they generally do not include effort data, which is essential to addressing most relevant questions in TD research. “Data” provided with published studies is often really the analysis of the raw data, not the data itself.
- benchmarks: Ground truth in this area is difficult to come by, but we can strive for intersubjectivity, or widespread agreement on an essentially subjective proposition. This can apply to, for example, tools that detect code smells or calculate code-based quality measures. Designating certain tools as “reference” implementations of these functions could also be useful.
- common metrics for outcome variables (maintenance effort, defects, etc.)
- tools: pluggable, validated, benchmarked. Tools must implement an underlying, understandable process, as some contexts will not allow for using a specific tool
- infrastructure for replication as well as for new work, so that it can be usefully compared to existing work

5 Open Problems

5.1 The Interplay Between Architectural (or Model Driven) Technical Debt vs. Code (or Implementation) Technical Debt

Clemente Izurieta (Montana State University – Bozeman, US)

License © Creative Commons BY 3.0 Unported license
© Clemente Izurieta

Joint work of Rojas, Gonzalo; Izurieta, Clemente

Main reference C. Izurieta, G. Rojas, I. Griffith, “Preemptive Management of Model Driven Technical Debt for Improving Software Quality”, in Proc. of the 11th Int’l ACM SIGSOFT Conf. on Quality of Software Architectures (QoSA’15), pp. 31–36, 2015; pre-print available from author’s webpage.

URL <http://dx.doi.org/10.1145/2737182.2737193>

URL <http://www.cs.montana.edu/izurieta/pubs/qosa39s.pdf>

The purpose of this open space session was to discuss the differences between architectural and model driven technical debt vs. code or implementation technical debt. Although terminology is used interchangeably, the community agrees that architectural debt occurs at a higher level of abstraction than implementation. Further, architectural debt encompasses tools that may be textual (i.e., RBML) or visual (i.e., UML, AADL), although the majority tends to be visually oriented. There is however a middle layer that overlaps between code-based implementation technical debt and architectural technical debt and it aggregates various metrics that can be measured at either level of abstraction with differing maintainability (architecture or code) consequences. This is an area that requires further research. Amongst other issues brought up was the technical debt associated with the automatic generation

of code from models. In some cases the code generated may be smelly because changes at the model level (that address a model smell) may generate constructs in the code that are deemed code smells. For example, empty classes and “to-do” tags. We also discussed the issue of traceability, and although a large community exists, its focus is on requirements; whereas traceability in this space refers to the interplay between technical debt observed in the code and its corresponding model or architecture. The latter may also lead to potential cause and effect studies.

5.2 From Technical Debt to Principal and Interest

Andreas Jedlitschka (Fraunhofer IESE – Kaiserslautern, DE), Liliana Guzmán, and Adam Trendowicz

License © Creative Commons BY 3.0 Unported license
© Andreas Jedlitschka, Liliana Guzmán, and Adam Trendowicz

Technical debt leads to interest payments in the form of additional effort that software practitioners need to do in future development of a software system because of quick and dirty design decisions. Deciding whether continuing paying interest or paying down the principal by refactoring the software is one of the most important challenges in managing technical debt. Recent experiences have shown that context-specific checklists to assess the implications of technical debt might be useful to support software practitioners in making this decision. However, software practitioners and researchers agreed on the need for context-specific estimation and prediction models regarding technical debt principal and interest. Furthermore, they claimed building such models involves the analysis of several factors including (1) the business goals and context associated to a software system, (2) the business implications associated to technical debt, (3) business as well as software development, maintenance, and refactoring costs, and (4) the complexity and benefits of the necessary changes.

5.3 Community Datasets and Benchmarks

Heiko Koziolk (ABB AG Forschungszentrum – Ladenburg, DE) and Mehdi Mirakhorli (Rochester Institute of Technology, US)

License © Creative Commons BY 3.0 Unported license
© Heiko Koziolk and Mehdi Mirakhorli

Research in managing technical debt could be improved if researchers shared information about their systems under analysis, datasets they produced, software tools they created and agreed on benchmarks to compare new methods and tools. In an Open Space Discussion, we brainstormed prerequisites and contents for a community repository that could contain these items. First, the technical debt community could appoint certain open source systems as community reference systems, which could be analyzed by different researchers. Besides the source code, also architectural knowledge, information about the development history, evolution scenarios and an issue tracker should be available for these systems, best facilitated by access to the system’s developers. Second, datasets could be shared in a similar fashion as in the PROMISE repository, for example anonymized analysis results from the company CAST or a set of reference technical debt metrics as proposed in context of the OMG’s Consortium

For IT Software Quality (CISQ) Seminar. Third, benchmarks could be established from these prerequisites, similar to the DEEBEE and BEFRIEND benchmarks used for evaluating design pattern detection and reverse engineering tools. A controlled experiment design could be another form of benchmark, allowing to evaluate the benefits of any approach managing technical in terms of quality and time-to-delivery. Finally, software tools, such as parsers, code smell detectors, pattern-detectors, issue tracker miners, etc., could be packaged and prepared for reuse by other researchers.

5.4 Deprecation

J. David Morgenthaler (Google Inc. – Mountain View, US)

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

Deprecation refers to a process for marking methods, classes, APIs, or entire systems whose use is to be discouraged. According to Wikipedia: While a deprecated software feature remains in the software, its use may raise warning messages recommending alternative practices; deprecated status may also indicate the feature will be removed in the future. Features are deprecated rather than immediately removed, to provide backward compatibility and give programmers time to bring affected code into compliance with the new standard.

The entire idea of software deprecation therefore maps directly to technical debt; it can be viewed as software aging made manifest. Deprecated features are the embodiment of an opening technology gap.

At Google, deprecation is often applied to discourage the use of features, but lacking discrete versions of most of our shared components, it is difficult to require users to upgrade their systems. Since everything is built from head, a feature cannot be deleted while existing users and their tests remain. Deprecation in this environment can become a tragedy of the commons. Yet our environment is closed, in that Google engineers can readily determine all users of a given feature. In fact, engineers are encouraged to find and update all uses of features they deprecate. As Google grows, however, this approach does not scale. As software everywhere become more dependent on services provided by the external environment, this type of technical debt will also become more prevalent. What can be done to avoid this onrushing crisis?

5.5 Report of the Open Space Group on Technical Debt in Product Lines

Klaus Schmid (Universität Hildesheim, DE), Andreas Jedlitschka (Fraunhofer IESE – Kaiserslautern, DE), John D. McGregor (Clemson University, US), and Carolyn Seaman (University of Maryland, Baltimore County, US)

License  Creative Commons BY 3.0 Unported license
© Klaus Schmid, Andreas Jedlitschka, John D. McGregor, and Carolyn Seaman

Product lines extend the traditional field of technical debt research as they widen the view towards whole sets of systems. With this wider view additional complexity comes in as relations among the systems may add to the debt. However, the first observation clearly is: all problems and issues of technical debt in single systems are relevant to product lines as well.

While, ideally, a product line is characterized by common reusable assets, often (code) clones may occur. This may happen either because the various products are not fully integrated, or because the development gets separated in an attempt to speed up development. This cloning can be a major source of technical debt. However, even if the product line is integrated and consists of a variability model and configurable assets, the chance of technical debt exists. Actually this situation may give rise to a particular form of technical debt: technical debt may exist due to (unnecessary) complexity created by variability models or variability markups in the code. Various indications may make such variability-related smells visible. These may either be described as metrics (e.g., total number of dependencies relative to features could be very high) or they can be described as logical anomalies (e.g., dependencies that indirectly lead to equivalences).

5.6 Automating Technical Debt Removal

Will Snipes (ABB – Raleigh, US) and Andreas Jedlitschka (Fraunhofer IESE – Kaiserslautern, DE)

License © Creative Commons BY 3.0 Unported license
© Will Snipes and Andreas Jedlitschka

Main reference N. Tsantalis, T. Chaikalis, A. Chatzigeorgiou. “JDeodorant: Identification and Removal of Type-Checking Bad Smells”, in Proc. of the 12th European Conf. on Software Maintenance and Reengineering (CSMR’08), pp. 329–331, IEEE CS, 2008.

URL <http://dx.doi.org/10.1109/CSMR.2008.4493342>

Organizations with legacy software may have technical debt in that software that they carry forward with each release. Although this debt can be identified, the manual process to identify and in particular remove debt items poses both challenges of economics and risk. Improvements are typically recommended only when changing the code for new features, because the changes pose risk to the system that must be mitigated with additional testing. One way to mitigate the risk and improve the effort required is to provide automated tool support for addressing technical debt items. Common tools that support automated refactoring cover only a few simple refactoring patterns that do not address more elaborate technical debt items such as code smells and code patterns. Recently, a couple of approaches have attempted more complex automation. DoctorQ by Dr. Jörg Rech is a plug-in for Eclipse that identifies the presence of anti-patterns while a developer works. Jdeodorant by Nikolaos Tsantalis is an Eclipse plugin that can automatically remove a few code smells, for example feature envy, from Java code. These tools provide a glimpse of what is possible with human-in-the-loop automation for addressing technical debt. Because of the risk of change previously mentioned, it is likely that any automation will require the developer to arbitrate between considerations to approve any proposed changes. Thus the automated technical debt challenge becomes to create automated solutions that address more complex technical debt patterns while the developer works in the code.

5.7 Social Debt in Software Engineering: Towards a Crisper Definition

Damian Andrew Tamburri (Polytechnic University of Milan, IT), Bill Curtis (CAST – Fort Worth, US), Steven D. Fraser (HP Inc. – Palo Alto, US), Alfredo Goldman (University of Sao Paulo, BR), Johannes Holvitie (University of Turku, FI), Fabio Queda Bueno da Silva (Federal University of Pernambuco – Recife, BR), and Will Snipes (ABB – Raleigh, US)

License © Creative Commons BY 3.0 Unported license
 © Damian Andrew Tamburri, Bill Curtis, Steven D. Fraser, Alfredo Goldman, Johannes Holvitie, Fabio Queda Bueno da Silva, and Will Snipes

Sustainable and scalable software systems require careful consideration over the force known as technical debt, i.e., the additional project cost connected to sub-optimal technical decisions. However, the friction that software systems can accumulate is not connected to technical decisions alone, but reflects also organizational, social, ontological and management decisions that refer to the social nature and any connected social debt of software – this nature is yet to be fully elaborated and understood. In a joint industry & academia panel, we refined our understanding of the emerging notion of social debt in pursuit of a crisper definition. We observed that social debt is not only a prime cause for technical debt but is also tightly knit to many of the dimensions that were observed so far concerning technical debt, for example software architectures and their reflection on organizations. Also we observed that social debt reflects and weighs heavily on the human process behind software engineering, since it is caused by circumstances such as cognitive distance, (lack of or too much of) communication, misaligned architectures and organizational structures.

The goal for social debt in the next few years of research should be to reach a crisp definition that contains the essential traits of social debt which can be refined into practical operationalizations for use by software engineering professionals in need of knowing more about their organizational structure and the properties/cost trade-off that structure currently reflects.

5.8 An Advanced Perspective for Engineering and Evolving Contemporary Software

Guilherme Horta Travassos (UFRJ / COPPE, BR)

License © Creative Commons BY 3.0 Unported license
 © Guilherme Horta Travassos
URL <http://www.cos.ufrj.br/~ght>

Software engineers continuously wonder about software and its nature. At the same time, software engineers should build high quality, low cost, on time and useful (software) products. However, the development of software does not follow a smooth path. The nature of software is more than technical (it is at least socio-technical), contributing to making the development process less predictable and prone to risks. Some issues, such as software defects, can be early identified, measured, and mitigated contributing to increasing its quality. Others, we lack evidence on their identification, measurement, and mitigation. In fact, software engineers are aware that some technical and socio-technical issues can jeopardize the software and its evolution. The decisions made by software engineers throughout the software life cycle are sources of technical and socio-technical issues. Such decisions can get a software project into “debt” and will affect the software project in the future. Some “debts” are invisible (software engineers can feel, but cannot see), others perceivable (software engineers can see,

but not feel), others concrete (software engineers can see and feel) and finally others are still intangible (no feel, no see, and something still happens in the project). All of them can influence positively or negatively the software project depending on the software ecosystem conditions. Such debts represent a promising perspective to deal with the quality and risks in contemporary software projects. The characterization and understanding of technical and no technical software debts represent an important step towards a better comprehension of the engineering process, the software nature, and its evolution.

Participants

- Areti Ampatzoglou
University of Groningen, NL
- Francesca Arcelli Fontana
University of Milano-Bicocca, IT
- Rami Bahsoon
University of Birmingham, GB
- Ayse Basar Bener
Ryerson Univ. – Toronto, CA
- Frank Buschmann
Siemens AG – München, DE
- Alexandros Chatzigeorgiou
University of Macedonia –
Thessaloniki, GR
- Zadia Codabux
Mississippi State University, US
- Bill Curtis
CAST – Fort Worth, US
- Steven D. Fraser
HP, Inc. – Palo Alto, US
- Alfredo Goldman
University of São Paulo, BR
- Christine Hofmeister
East Stroudsburg University, US
- Johannes Holvitie
University of Turku, FI
- Clemente Izurieta
Montana State University –
Bozeman, US
- Andreas Jedlitschka
Fraunhofer IESE –
Kaiserslautern, DE
- Sven Johann
innoQ GmbH – Monheim am
Rhein, DE
- Heiko Koziolok
ABB AG Forschungszentrum –
Ladenburg, DE
- Philippe Kruchten
University of British Columbia –
Vancouver, CA
- Jean-Louis Letouzey
inspearit – Paris, FR
- Antonio Martini
Chalmers UT – Göteborg, SE
- John D. McGregor
Clemson University, US
- Mehdi Mirakhorli
Rochester Institute of
Technology, US
- J. David Morgenthaler
Google, Inc. –
Mountain View, US
- Robert Nord
Carnegie Mellon University –
Pittsburgh, US
- Ipek Ozkaya
Carnegie Mellon University –
Pittsburgh, US
- Fabio Queda Bueno da Silva
Federal University of
Pernambuco – Recife, BR
- Klaus Schmid
Universität Hildesheim, DE
- Carolyn Seaman
University of Maryland,
Baltimore County, US
- Andriy Shapochka
SoftServe – Lviv, UA
- Forrest Shull
Carnegie Mellon University –
Pittsburgh, US
- Will Snipes
ABB – Raleigh, US
- Damian Andrew Tamburri
Polytechnic Univ. of Milan, IT
- Graziela Tonin
University of São Paulo, BR
- Guilherme Horta Travassos
UFRJ / COPPE, BR



Algorithmic Methods for Optimization in Public Transport

Edited by

Leo G. Kroon¹, Anita Schöbel², and Dorothea Wagner³

¹ Erasmus University – Rotterdam, NL, lkroon@rsm.nl

² Universität Göttingen, DE, schoebel@math.uni-goettingen.de

³ KIT – Karlsruher Institut für Technologie, DE, dorothea.wagner@kit.edu

Abstract

This report documents the talks and discussions at the Dagstuhl seminar 16171 “Algorithmic Methods for Optimization in Public Transport”. The seminar brought together researchers from algorithm, algorithm engineering, operations research, mathematical optimization and engineering, all interested in algorithms in public transportation. Also several practitioners were able to join the group and brought valuable insights on current practice and challenging problems.

Seminar April 24–29, 2016 – <http://www.dagstuhl.de/16171>

1998 ACM Subject Classification G.1.6 Optimization, G.2.1 Combinatorics, G.2.2 Graph Theory, G.2.3 Applications

Keywords and phrases delay and disruption management, dynamic passenger information, public transportation, resource scheduling, timetabling

Digital Object Identifier 10.4230/DagRep.6.4.139

Edited in cooperation with Jonas Harbering

1 Executive Summary

Leo G. Kroon

Anita Schöbel

Dorothea Wagner

License © Creative Commons BY 3.0 Unported license
© Leo G. Kroon, Anita Schöbel, and Dorothea Wagner

Public transport systems are highly complex systems, due to their technical and organizational complexity, and due to the large numbers of passengers that are transported each day. The quality of the services provided to the passengers is on the one hand the result of the quality and robustness of the underlying plans, such as the timetable and the vehicle and crew schedules. On the other hand, in real-time the quality of the service is the result of the complex interactions between the real-time logistic management of the public transport system and the information to and guidance of the passengers.

Both in the planning stage and in real-time, dealing with these problems requires handling large amounts of data, solving complex combinatorial optimization problems, and dealing with uncertainty. Preferably, the optimization models aim to improve the robustness of the public transport system, so that the system is less vulnerable to disturbances.

In addition, due to the use of smart cards and smart phones, it becomes technically possible to give personalized real-time traffic advice for passengers to guide them to their destinations, even in disturbed situations. In addition, the use of these devices makes huge amount of data available, which can improve decisions in real-time control and in disruption management as well as in the planning stage.



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

Algorithmic Methods for Optimization in Public Transport, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 139–160

Editors: Leo G. Kroon, Anita Schöbel, and Dorothea Wagner



Dagstuhl Reports

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

In this seminar, researchers from algorithm engineering and operations research worked together with researchers with an engineering background and participants from practice. The common goal was to improve methods for planning and scheduling of public transportation.

Among others, some specific topic which were covered were

- *Scheduling* of public transport. Several new applications and new ideas on algorithms for public transport scheduling were presented.
- *Integration of planning stages*. Suggestions were developed on how the traditional approach of sequential planning can be replaced by integrated approaches.
- *Robustness and recoverability*. Here several talks discussed methods on how to react to different kinds of disturbances, or how to make schedules more robust.
- *Real-time control*. Real-time control measures which can be taken to get back to the plan as soon as possible were proposed and discussed.
- *Routing in public transport*. For the important issue of routing passengers in public transport, also needed for timetable information systems, several algorithms and new approaches were presented and discussed.
- *Applications and case studies*. Among others the situation in Mumbai, India, was presented and discussed and representatives of several public transport operators sketched the planning process in their companies and pointed out open questions for further research.
- *Future technologies* were another important issue. The participants discussed the potential of new technologies and identified algorithmic challenges for their future utilization.

The seminar started with an introductory round in which every participant presented him- or herself with three slides. It was a good start to get to know each other. In the following days, nearly all participants contributed with talks. There were also two panel discussions, one with the other Dagstuhl group on learning algorithms, and another one on future technologies. The participants discussed and identified challenging algorithmic problems in this field.

Finally, the organizers would like to thank the Dagstuhl team and all the participants for a fruitful and successful seminar.

Leo Kroon, the main organizer of this Dagstuhl seminar, died unexpectedly on 14 September 2016. We are shocked and very sad about his sudden death. Leo was a great scientist and a wonderful person. We will never forget him.

2 Table of Contents

Executive Summary

Leo G. Kroon, Anita Schöbel, and Dorothea Wagner 139

Overview of Talks

Integrated Public Transport Optimisation and Planning <i>Allan Larsen</i>	143
Hypergraphs in traffic optimization <i>Ralf Borndörfer</i>	143
Timetable Planning of a High-Speed Chinese Railway Corridor <i>Valentina Cacchiani</i>	143
Railway traffic control targeting passenger flows <i>Francesco Corman</i>	144
Industrial issues for Mass Transit Operations <i>David De Almeida</i>	144
Challenges in Public Transport and in Public Transport Planning <i>Markus Friedrich</i>	145
Public Transport in Emergency Planning <i>Marc Goerigk</i>	146
Dealing with uncertainty in railway traffic management and disruption management <i>Rob Goverde</i>	146
Further Insight into Single Track Train Scheduling <i>Jonas Harbering</i>	147
Hyperpaths for Public Transport Assignment <i>Mark Hickman</i>	147
Computing and Improving Passenger Punctuality <i>Dennis Huisman</i>	148
Smart Route Planning for Public Transit <i>Giuseppe F. Italiano</i>	148
Robust Efficiency for Resource Scheduling: Insights from Public Bus Transport and Airline Cases <i>Natalia Kliewer</i>	148
Challenges for Public Transport <i>Leo G. Kroon</i>	149
Integrating Passenger Assignment and Timetabling for Capacitated Public Transit Networks <i>Marco Laumanns</i>	149
Real-time Re-scheduling for Public Transit <i>Janny Leung</i>	150
Integrated railway operations planning: What we can do and how? <i>Lingyun Meng</i>	150

A non-compact formulation for job-shop scheduling problems in transportation <i>Carlo Mannino</i>	151
A convex programming approach for stochastic timetable optimisation <i>Gabor Maroti</i>	151
Objectives in Rapid Transit Network Design and Line Planning <i>Juan Antonio Mesa and Francisco A. Ortega</i>	152
Robust Routing in Urban Public Transportation: Evaluating Strategies that Learn From the Past <i>Matus Mihalak</i>	153
MIDAS-CPS: A Possible Future of Proactive Traffic Management Systems <i>Pitu Mirchandani</i>	153
Rolling stock planning and challenges, DSB, Denmark <i>Morten Nyhave Nielsen</i>	154
PANDA – A Framework for Passenger-Oriented Train Disposition <i>Matthias Müller-Hannemann</i>	154
Mathematical Modelling of Industrialized Timetables <i>Karl Nachtigall</i>	155
O.R. for conventional rail operations in India <i>Narayan Rangaraj</i>	155
Train routing selection for the real-time railway traffic management <i>Marcella Sama, Andrea D’Ariano, Dario Pacciarelli, Paola Pellegrini, and Joaquin Rodriguez</i>	156
Approaches for integrated planning in public transportation <i>Anita Schöbel</i>	156
From Robustness to Resilience – How to evaluate resilience of a timetable <i>Norio Tomii</i>	157
Decision-support for railway traffic management: What can we actually conclude from previous research work and does the work meet the needs experienced by practitioners? <i>Johanna Törnquist Krasemann</i>	157
The future of public transport by autonomous vehicles <i>Lucas Veelenturf and Afonso Sampaio</i>	158
Engineering Graph-Based Models for Dynamic Timetable Information Systems <i>Christos Zaroliagis</i>	158
Participants	160

3 Overview of Talks

3.1 Integrated Public Transport Optimisation and Planning

Allan Larsen (*Technical University of Denmark – Lyngby, DK*)

License © Creative Commons BY 3.0 Unported license
© Allan Larsen

Joint work of Stefan Ropke, Roberto Roberti, Evelien van der Hurk, Joao Fonseca, Allan Larsen

The planning of efficient and cost-effective public transport systems offering high-class service to the users is a very complex task. The Integrated Public Transport Optimisation and Planning (IPTOP) research project aims at developing new methods for planning, designing, controlling and optimising Danish public transport systems. The IPTOP project financed by the Strategic Research Council of Denmark spans over a five year period and has just entered into its second year. In this presentation I will provide an overview of the overall project aims as well as introduce our current efforts of integrating some of subproblems of the planning process into optimisation models. These works include a project on integration of bus timetabling with vehicle scheduling and another project on integration of train timetabling and rolling stock planning.

3.2 Hypergraphs in traffic optimization

Ralf Borndörfer (*Konrad-Zuse-Zentrum – Berlin, DE*)

License © Creative Commons BY 3.0 Unported license
© Ralf Borndörfer

Joint work of Marika Karbstein, Isabel Beckenbach, Olga Heismann, Markus Reuther, Thomas Schlechte, Christof Schulz, Steffen Weider, Ralf Borndörfer

URL <http://www.zib.de/projects/service-design-public-transport>

URL <http://www.zib.de/projects/modal-raillab>

Traffic optimization is intimately related to algorithmic graph theory, which provides elegant solutions to problems ranging from network design to vehicle rotation planning. Extending these approaches to a hypergraph setting is a natural next step that allows to deal, in a mathematically appealing way, with complex types of constraints beyond the node-edge level. The talk illustrates the potential of hypergraph models on two examples in line planning and railway vehicle rotation planning. Line planning gives rise to the Steiner path connectivity problem, a generalization of the Steiner tree problems to hypergraphs, while railway vehicle rotation planning leads to the consideration of hyperassignment problems. These models, their theory, and algorithmic solution will be discussed.

3.3 Timetable Planning of a High-Speed Chinese Railway Corridor

Valentina Cacchiani (*University of Bologna, IT*)

License © Creative Commons BY 3.0 Unported license
© Valentina Cacchiani

Joint work of Feng Jiang, Paolo Toth, Valentina Cacchiani

In this work, we consider the Train Timetabling Problem (TTP) in the planning phase, which calls for determining an optimal schedule for a given set of trains, while satisfying

several constraints related to safety and to the physical infrastructure. In particular, we focus on the TTP of the high-speed trains of the Chinese Beijing-Shanghai corridor. Two are the objectives to be achieved, namely scheduling as many trains as possible and obtaining a regular schedule, i.e. a schedule showing regularity in the train frequency at the main stations. We present a Lagrangian-based heuristic algorithm adapted from a previous work to deal with new objectives and constraints. In addition, we present our research ideas on this problem and on the general TTP.

3.4 Railway traffic control targeting passenger flows

Francesco Corman (TU Delft, NL)

License © Creative Commons BY 3.0 Unported license
© Francesco Corman

Joint work of Dario Pacciarelli, Alessio D. Marra, Andrea D’Ariano, Marcella Samà, Francesco Corman

Main reference F. Corman, A. D’Ariano, A. D. Marra, D. Pacciarelli, M. Samà, “Integrating train scheduling and delay management in real-time railway traffic control,” *Transportation Research Part E: Logistics and Transportation Review*, 2016.

URL <http://dx.doi.org/10.1016/j.tre.2016.04.007>

Optimization models for railway traffic rescheduling tackle the problem of determining, in real-time, control actions reducing effect of disturbances in railway systems. In this field, mainly two research streams can be identified. On the one hand, train scheduling models are designed to include all conditions relevant to feasible and efficient operation of rail services, from the viewpoint of operations managers. On the other hand, delay management models focus on the impact of rescheduling decisions on the quality of service perceived by the passengers. Models in the first stream are mainly microscopic, while models in the second stream are mainly macroscopic. This paper aims at merging these two streams of research by developing microscopic passenger-centric models, solution algorithms and lower bounds. Several fast heuristic methods are proposed, based on alternative decompositions of the model. A lower bound is proposed, consisting of the resolution of a set of min-cost flow problems with activation constraints. Computational experiments, based on multiple test cases of the real-world Dutch railway network, show that good quality solutions and lower bounds can be found within a limited computation time.

3.5 Industrial issues for Mass Transit Operations

David De Almeida (SNCF – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© David De Almeida

This talk will focus on public transport systems in their Mass Transit version and will be illustrated with the context of the Paris region. It will underline the major role of heavy rail, together with other multimodal complements. After reviewing some facts about the volume of customers (in significant growth those last 10 years) and trains that need to be matched, some issues related to the transport capacity and customers service quality will be discussed. The talk will also illustrate short to long term opportunities for decision-making and OR models from the point of view of the Innovation & Research Division of SNCF. In particular, we will underline needs for research and approaches enhancing operation monitoring, disturbance

early detection and real-time decision aid. Models for integrating the planning and real-time management of operations in complex local areas (like major stations where timetabling, platforming and shunting problems must be simultaneously dealt with) will also be discussed. We will finally share some thoughts about requirements (data access, human factors...) that need to be addressed in order to leverage the actual use of such works in railway companies.

3.6 Challenges in Public Transport and in Public Transport Planning

Markus Friedrich (*Universität Stuttgart, DE*)

License  Creative Commons BY 3.0 Unported license
© Markus Friedrich

The presentation summarizes various challenges public transport may face in the coming decades.

1. Demand side challenges.
Will demand for public transport increase or decrease? Fewer younger people and a higher share of wealthy elderly may increase car ownership and may lead to less demand for public transport in urban areas. Young adults which seem to be more open to sharing concepts may increase demand for public transport. As expanding companies more and more move out of cities the share of public transport for work related trips may decrease. Forecasts for the German Transport Master Plan indicate that local public transport will increase less than car transport and long distance public transport.
2. Supply side challenges.
Can public transport compete with the car? Currently the mode car serves 80% of of all person-kilometres travelled in Germany. One reason for this is simply that car is faster. Advantages of public transport result from lower travel costs, from shorter travel times to dense urban centres during peak hours and from the possibility to use the travel time for reading or other activities. The next generation of cars will consume less energy, will produce fewer local emissions, will be more comfortable and always connected, will assist the driver, will drive partly autonomously and park autonomously. This will lead to higher traffic safety and more benefits for car-users as they can perform other activities during their journey. As a consequence, it is likely that public transport will lose travellers to public transport. If society does not like this development, transport policy should introduce road pricing schemes and use the revenues to improve public transport.
3. Challenges in planning.
What are obstacles in public transport planning? Major obstacles result from the fact that transport planning (planning of stops, line routes and headways) and operation planning (timetable planning, vehicle and driver scheduling) are not integrated.
4. Challenges in operation.
How can we handle disturbances?
Public transport networks are more vulnerable to network disturbances than car networks. Providing redundancy is expensive. For the daily operation standard fleet management systems still provide little support for handling disturbances. To overcome these shortcomings, it is necessary to develop and implement methods for better forecasting the downstream travel times of public transport vehicles. Better vehicle dispatching systems should support dispatchers by suggesting suitable measures.

5. Challenges in financing.

How can we finance future public transport?

In urban public transport in Germany passengers pay roughly 1€/trip or 2€/day. Electronic ticketing still is very basic and not customer friendly. Best pricing options are not available. Bus drivers earn approx. 15€/h, need to work extra hours to feed a family, are hard to find in some areas and therefore should not be the main source when minimizing costs. Cost coverage of operating costs (labour, rolling stock, fuel) lies between 50% to 70%. Additional supply usually increases the deficit. Introduction of smaller vehicles can provide more economic solutions than stand vehicle sizes. As small vehicles also require drivers they can never provide competitive prices and require high subsidies per traveller. This may change with self-driving busses, which will probably be only available after the self-driving car. Financing local public transport will therefore continue to require public subsidies. This can be justified by the positive impacts of public transport to society. Public transport reduces congestion, provides travel opportunities for everybody and is more environmental friendly than cars.

6. Conclusion.

Public Transport is great as it provides mobility options to everybody, avoids auto-dependency, is the best ride-sharing system available, is the only system which can provide high capacity transport for metropolitan areas even with self-driving cars, needs less energy per person kilometres than other modes, if occupancy rate is > 40%. Public Transport has shortcomings as it is only faster than car in case of high speed trains or in congested urban areas, does often not offer competitive prices, is less comfortable than a car, is often not reliable and requires subsidies, which must be justified.

3.7 Public Transport in Emergency Planning

Marc Goerigk (Lancaster University, GB)

License  Creative Commons BY 3.0 Unported license
© Marc Goerigk

Thinking of an emergency evacuation, pictures that come first to our mind would be cars stuck in long queues, or masses of pedestrians. However, public transport plays its important, but relatively little known part in emergency planning. The purpose of this talk is to give an overview of the applications and research fields involved, to promote a closer collaboration between the two research communities, and to point out current research questions.

3.8 Dealing with uncertainty in railway traffic management and disruption management

Rob Goverde (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Rob Goverde

Uncertainty is an essential part of railway operations. In disruption management the durations of disruptions is uncertain and estimations typically have large variance. Likewise, in traffic management conflict predictions have some uncertainty although with less variance. How to

deal with this uncertainty to get high performance of operations control? This is the focus of this talk.

3.9 Further Insight into Single Track Train Scheduling

Jonas Harbering (Universität Göttingen, DE)

License  Creative Commons BY 3.0 Unported license
© Jonas Harbering

Joint work of Jonas Harbering, Abhiram Ranade, Marie Schmidt, Oliver Sinnen

In this talk we present the single track train scheduling problem. The idea for studying this problem was raised by the current situation of public transportation in India. Large parts of the infrastructure network are composed of single track stretches and the aim is to use the given capacitated infrastructure as good as possible.

In the single track train scheduling problem a linear network is given. While the capacity of the track segments is one, meaning between two consecutive stations there can only be one train at any given time, the capacity of the stations is unlimited. A set of trains from the left hand side and a set of trains from the right hand side are given which are to traverse the entire network towards the right hand side and left hand side, respectively. The aim is to minimize the makespan, i.e., the time from the first departure of the first train until the last arrival of the last train.

Some results on polynomial and weakly polynomial time algorithms for instances with only four stations are shown to solve the problem. Furthermore, a dynamic programming algorithm is shown to have a pseudo-polynomial runtime, given that the number of stations is fixed. Finally, extensions for this problem leading to more realistic settings are discussed.

References

- 1 J. Harbering, A. Ranade, M. Schmidt, O. Sinnen *Single Track Train Scheduling*. Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen

3.10 Hyperpaths for Public Transport Assignment

Mark Hickman (The University of Queensland, AU)

License  Creative Commons BY 3.0 Unported license
© Mark Hickman

Passenger assignment involves determining passenger flows in a public transit network. These assignment models are typically associated with individual passenger behaviors, such as a “strategy” among attractive lines at a stop or a discrete choice among scheduled services. In these cases, assignment can be described using hyperpaths. From empirical data, such hyperpaths can be observed in practice. However, identifying these hyperpaths is computationally challenging, and calibrating passenger behavioral models is similarly complex. These challenges are discussed, along with both algorithmic successes and limitations for these hyperpaths.

3.11 Computing and Improving Passenger Punctuality

Dennis Huisman (Erasmus University – Rotterdam, NL)

License  Creative Commons BY 3.0 Unported license
© Dennis Huisman

In this talk, we present new key performance indicators (KPIs) on passenger punctuality. These KPIs are part of the contract between the Dutch government on one side, and NS and ProRail on the other side.

In the presentation, we will explain how the KPIs are computed. Moreover, we would like to have an discussion on ideas how these KPIs can be improved.

3.12 Smart Route Planning for Public Transit

Giuseppe F. Italiano (University of Rome “Tor Vergata”, IT)

License  Creative Commons BY 3.0 Unported license
© Giuseppe F. Italiano

Joint work of Luca Allulli, Donatella Firmani, Luigi Laura, Federico Santaroni, Giuseppe F. Italiano

Current journey planners for public transport are mostly based on timetabling information only, i.e., they hinge on the assumption that all transit vehicles run on schedule. Unfortunately, this might not always be realistic, as unpredictable delays may occur quite often in practice. In this case, it seems quite natural to ask whether the availability of real-time updates on the geo-location of transit vehicles may help improving the quality of the solutions offered by routing algorithms. To address this question, we considered the public transport network of the metropolitan area of Rome, where delays are not rare events, and reported the results of our experiments with two journey planners that are widely used for this city: one based on timetabling information only (Google Transit) and one which makes explicit use of GPS data on the geo-location of transit vehicles (Muovi Roma).

3.13 Robust Efficiency for Resource Scheduling: Insights from Public Bus Transport and Airline Cases

Natalia Kliewer (FU Berlin, DE)

License  Creative Commons BY 3.0 Unported license
© Natalia Kliewer

Joint work of Natalia Kliewer, Bastian Amberg, Lucian Ionescu

We analyze and compare robust efficiency of resource schedules in public bus transportation and airline industry, both dealing with the competing objectives of cost-efficiency and schedule robustness under delays. Generalizing the findings from robust and efficient vehicle and crew scheduling in both airline and public transport contexts, we compare different techniques that lead to an improvement of the non-dominated solution front.

3.14 Challenges for Public Transport

Leo G. Kroon (Erasmus University – Rotterdam, NL)

License  Creative Commons BY 3.0 Unported license
© Leo G. Kroon

We describe several challenges for the further improvement of public transport systems. These challenges are partially related to further improving the current supply-oriented public transport systems as they are currently, usually based on a fixed line system and timetable. As has been described in other abstracts, these challenges are in the following areas: passenger orientation, integration of planning stages, robustness and resilience, energy efficiency, and disruption management. Another challenge is to effectively use the wealth of big data that is available nowadays to improve the quality of public transport systems. The challenges are also partially related to future demand oriented public transport systems that may be operated in a number of years from now. These systems may be operated with electrical autonomous cars, and provide more personalized public transport. Challenges in this area may be related to determining how many vehicles are needed to operate the system at a certain service level, which dispatching and routing strategies to use in real-time for routing the vehicles, how to route the passengers through the system, which charging strategies to use to guarantee that the vehicles do not run out of power, and which pricing strategies to use to manage the passenger demand? Such questions provide an interesting new research agenda for the public transport research community.

3.15 Integrating Passenger Assignment and Timetabling for Capacitated Public Transit Networks

Marco Laumanns (IBM Research Zurich, CH)

License  Creative Commons BY 3.0 Unported license
© Marco Laumanns
Joint work of Marco Laumanns, Jacint Szabo, Maya Voegeli

In this talk we present a bilevel optimization model for the integration of passenger assignment into the (periodic) timetabling problem for capacitated public transportation networks. For the lower level problem we present a mixed-integer problem formulation which is based on the assumption that passengers are daily commuters with perfect information and which takes into account selfish routing and prioritization of already on-board passengers over boarding passengers. The integration of this problem formulation into the timetabling problem results in a mixed-integer bilinear programming problem. In order to solve this problem we propose a heuristic solution approach for general instances. Additionally, we present a direct approach which is obtained by a mixed-integer linear program reformulation using unary or binary expansion and McCormick relaxation. To improve the performance of the direct approach we provide problem-specific cutting planes as well as a reduction of the number of binary variables motivated by analysis of the problem structure.

3.16 Real-time Re-scheduling for Public Transit

Janny Leung (The Chinese University of Hong Kong, HK)

License © Creative Commons BY 3.0 Unported license
© Janny Leung

Joint work of KUO Yong-Hong, LAI Shuwo David, CHEUNG Kam Fung Henry
Main reference J. M. Y. Leung, D. S. W. Lai, Y.-H. Kuo, H. K. F. Cheung, “Real-Time Integrated Re-scheduling for Public Transit”, in Proc. of the Third Int’l Conf. on Railway Technology: Research, Development and Maintenance, Civil-Comp Press, 2016.

URL <http://dx.doi.org/10.4203/ccp.110.287>

We study a vehicle and crew (re-)scheduling problem for a public transit system which is subject to highly stochastic travel times and disruptions. Our research is motivated by the operations of the tramways system in Hong Kong, which serves hundreds of thousands of passengers per day in a densely populated area and whose operations face severe challenges because it does not run on dedicated tracks but must share the road with vehicular traffic in heavily congested areas.

We investigate how the availability of historical and real-time location and traffic information can be exploited to provide decision support to the controllers. We develop a model for the re-scheduling of vehicles and crew, so as to maximize the route frequencies in order to provide good service to passengers, and minimize the violation of staff regulations (meal-break delays and overtime) taking stochastic time-dependent travel-time uncertainties into account.

In the operations that motivated our research, re-assignment of motormen/trams to different routes is possible only upon arrival at a terminal. Therefore, our decision support system is a “look-ahead” model (solved repeatedly on a rolling horizon basis) to find the best set of re-assignments for all trams and motormen that will be arriving at some terminal within the next time period. For all trams arriving at a terminal within the planning period, we consider all possible subsequent schedules that could be assigned to the tram and evaluate the cost (in terms of overtime, demand coverage, meal-break delays, etc.). Using a matching-based model, re-assignment of routes for all trams arriving at the terminals within the planning period is optimised.

We also explored a variant of the model where we consider demand constraints not only for the current planning period, but also several time periods into the future. Another variant of the model considers crew and tram availability. Yet another variant is the incorporation of planned maintenance of the vehicles into the daily scheduling. We are also interested in exploring the robustness of the model when the frequency of re-optimisation is increased. Future research will investigate how real-time demand information (from multi-media sources) might be available and how the system can operate to be more demand-responsive dynamically.

3.17 Integrated railway operations planning: What we can do and how?

Lingyun Meng (Beijing Jiaotong University, CN)

License © Creative Commons BY 3.0 Unported license
© Lingyun Meng

There is a trend that we integrate railway operations planning processes to get better solutions. This talk gives questions regarding necessity of integration, possible interesting and valuable integration related topics, challenges and potential methods to deal with them. Also we

report our ongoing work with regard to demand-oriented vehicle routing scheduling for a transport system and integrated railway traffic control and train control.

3.18 A non-compact formulation for job-shop scheduling problems in transportation

Carlo Mannino (SINTEF ICT – Oslo, NO)

License © Creative Commons BY 3.0 Unported license
© Carlo Mannino

Joint work of Leonardo Lamorgese, Carlo Mannino

A central problem in transportation is that of routing and scheduling the movements of vehicles so as to minimize the cost of the schedule. It arises, for instance, in timetabling, dispatching, delay and disruption management, runway scheduling, and many more. For fixed routing, the problem boils down to finding a minimum cost conflict-free schedule, i.e. a schedule where potential conflicts are prevented by a correct timing of the vehicles on the shared resources. A classical mathematical representation involves continuous variables representing times, (time-precedence) linear constraints associated with single vehicles, and disjunctive (precedence) linear constraints associated with pairs vehicles. There are two standard ways to linearize disjunctions, namely by means of Big_M formulations or by time-indexed formulations. Big_M formulations tend to return notoriously weak relaxations, whereas time-indexed formulations quickly become too large for instances of some practical interest. In this work we develop a new, non-compact formulation for such disjunctive programs with convex piece-wise linear cost, and solve the resulting problems by row-generation. Our initial tests show that the new approach favourably compares with the so-far most effective approach on a large number of real-life test instances from railway traffic management. Moreover, it opens up for several research directions, ranging from investigating polyhedral properties to algorithmic speed-ups.

3.19 A convex programming approach for stochastic timetable optimisation

Gabor Maroti (VU University Amsterdam, NL)

License © Creative Commons BY 3.0 Unported license
© Gabor Maroti

The punctuality of the trains has always been one of the most scrutinised, and often ridiculed, performance indicators of a railway network. Recent research has proposed various optimisation model for improving the delay absorption capacity of a timetable.

Kroon et al. [1] described a stochastic programming model. While the model has shown its value in practice, the solution approach does not scale well, and becomes barely tractable for practically interesting instances.

In this talk we propose a convex programming solution approach for the model of Kroon et al. [1]. It turns out that our algorithm solve large real-life instances very quickly and reliably, with tight optimality guarantees.

One can argue, though, that we are solving the wrong problem. We optimise the train delays and not the passenger delays. The next challenge is thus to understand to way how

passenger delays arise from train delays, and to incorporate this effect in the timetable optimisation algorithm.

References

- 1 L. G. Kroon, G. Maroti, M. Retel Helmrich, M. J. C. M. Vromans, and R. Dekker. Stochastic improvement of cyclic railway timetables. *Transportation Research Part B*, 42(6):553–570, 2008.

3.20 Objectives in Rapid Transit Network Design and Line Planning

Juan Antonio Mesa (University of Sevilla, ES) and Francisco A. Ortega

License © Creative Commons BY 3.0 Unported license
© Juan Antonio Mesa and Francisco A. Ortega

Planning rapid transit systems includes decisions about the design of the lines as well as frequency and capacity of the service. These decisions seek a trade-off between supply and demand. There are two main kinds of agents interested in the network design and line planning process: travelers and construction and operating companies. Moreover, local, regional or state authorities represented by the transportation agency are also a part. The general aim of a rapid transit system is to improve the mobility of citizens. One of the aims of the authorities is to facilitate the accessibility to rapid transit and to maximize the coverage. Travelers want to have the possibility of traveling at any time, as fast and as cheaply as possible, and without transfers. Companies want to minimize construction, vehicle, and operating costs, and maximize revenue. Therefore, there are many functions to be applied as objectives in network design and line planning. Among design variables are node and edge selection, frequency and capacity of the trains. Most of the scientific literature is dedicated to studying a particular problem with one objective function. However, several questions arise when dealing with a real-life problem: which is the measure to be used as objective function? Are there analytical or experimental relationships between identical problems with different objective functions? How to weigh linear combinations of different measures? What about globalizing function and multicriteria approaches? Although most of the problems are NP-hard, are there significant differences in computational times? In this talk an overview of the scientific challenges that presents this kind of problems has been presented.

References

- 1 David Canca, Alicia de-Los-Santos, Gilbert Laporte, Juan A. Mesa *A general rapid transit network design, line planning and fleet investment integrated model*. *Annals of Operations Research*, DOI 10.1007/s10479-014-1725.0, 2014.
- 2 R. van Nes, P.H.L. Bovy *Importance of Objectives in Urban Transit-Network Design*. *Transportation Research Record* 1735, 25–34, 2000.
- 3 Sutapa Samanta, Manoj K. Jha *Modeling a rail transit alignment considering different objective* *Transportation Research, Part A*, 45, 31–45, 2011.
- 4 Anita Schöbel *Line Planning in Public Transportation: Models and Methods* *OR Spectrum*, 34, 491–510, 2012.

3.21 Robust Routing in Urban Public Transportation: Evaluating Strategies that Learn From the Past

Matus Mihalak (Maastricht University, NL)

License  Creative Commons BY 3.0 Unported license
© Matus Mihalak

Trams and buses of an urban transportation do not always run according to the timetable, because of unavoidable delays. Using journeys planned according to the timetable may be sub-optimal, if a carefully planned transfer fails because of a late arrival at the transfer bus stop. Robust routing aims at providing journeys in an urban transportation network that are (somewhat) resilient against delays. In this talk I present an approach to robust routing that learns from the past. In particular, we have access to the exact travel times of each of the vehicles in the system over the few past days, say, of the last 2-3 weeks. For every day, these travel times define a so-called recorded timetable (a timetable for which the buses and trams were not delayed on that day). We concretely investigate the following optimization problem: Given a target time T , an origin stop O and a destination stop D , what is the time-optimal journey that takes me from stop O to stop D on time, i.e., before time T ?

We investigate several heuristics for this robust routing question: (1) we adapt the min max relative regret approach that takes 2,3, or 7 past days as the reference set of instances (recorded timetables) for which we want to find a good journey, (2) we find a journey J that minimizes $\text{AVG}(J) + \alpha \cdot \text{DEVIATION}(J)$, where $\text{AVG}(J)$ is the average travel time of journey J over the recorded timetables, $\text{DEVIATION}(J)$ is the standard deviation of journey J , and α is a parameter (that can be tuned), (3) we compute an optimal journey that allows at least T minutes for transfer between two bus/tram lines. We use some of the recorded timetables for learning purposes, and some for testing. Our experimental results show that the approach (2) gives superior results. The approach, however, needs a fine tuning of α according to the recorded timetables. On the other hand, approach (1) does not require any fine-tuning at all, and is in every aspect very competitive with approach (2). Finally, we observe that approach (3) fails by far to compete with (1) and (2).

3.22 MIDAS-CPS: A Possible Future of Proactive Traffic Management Systems

Pitu Mirchandani (Arizona State University – Tempe, US)

License  Creative Commons BY 3.0 Unported license
© Pitu Mirchandani

While driving on your favorite route to your destination, have you ever wondered why the technology you are seeing as far as traffic management is concerned is so antiquated? My answer to that is the people and organization that manage the traffic are not “cyber-physicists” nor “real-time optimizers”. MIDAS hopes to demonstrate the synergistic use of a cyber-physical infrastructure consisting of smart-phone type devices; cloud computing, wireless communication, and intelligent transportation systems to manage vehicles in the complex urban network – through the use of traffic controls, route advisories, road pricing/rewards and route guidance – to jointly optimize drivers’/travelers mobility as well as achieve the sustainability goals of reducing energy usage and improving air quality. A key element of MIDAS-CPS is the real-time streaming data collection and data analysis and the subsequent

traffic management through proactive traffic controls and advisories, through visualizations of predicted queues ahead, effective road prices/rewards, and route advisories. Although drivers will not be forced to use recommended routes, it is anticipated that MIDAS-CPS would lead to lesser drive stress and improved road safety, besides the designed benefits on the environment, energy consumption, congestion mitigation, and driver mobility. This talk will only focus on overall architecture of MIDAS and on the proactive traffic management component, while the on-going sponsored multidisciplinary NSF project is at the cutting edge in several areas: real-time image processing, real-time traffic prediction and supply/demand management, and data processing/management through cloud computing.

3.23 Rolling stock planning and challenges, DSB, Denmark

Morten Nyhave Nielsen (DSB – Copenhagen, DK)

License  Creative Commons BY 3.0 Unported license
© Morten Nyhave Nielsen

The talk is divided into two parts. First, we consider the situation for DSB (Danish State Railways) in Denmark after a new contract was signed with the government in 2015, and second we look at how optimization is used in the planning process. In the contract of 2015, the punctuality measure has changed from train punctuality to customer punctuality. Besides improving punctuality, the government also has a vision of a “one hour model” with travel times below one hour between the larger cities in Denmark. This requires huge investments in infrastructure in the years to come. Optimization software is used both for rolling stock and crew at DSB. We experience a rather large gap between developing the models/algorithms and using the theory in practice. Besides coping with all the practical details, the solutions should include features to be accepted easier by the organization and the daily planners. i.e. plan should look similar between days. Many plan starts from an existing plan, we therefore encourage the community to increase the focus on “repairing” plans instead of “generating” plans.

3.24 PANDA – A Framework for Passenger-Oriented Train Disposition

Matthias Müller-Hannemann (Martin-Luther-Universität Halle-Wittenberg, DE)

License  Creative Commons BY 3.0 Unported license
© Matthias Müller-Hannemann

Joint work of Christoph Blendinger, Martin Lemnian, Steffen Rechner, Ralf Rückert

We introduce the decision support tool PANDA (Passenger Aware Novel Dispatching Assistance). Our web-based tool is designed to provide train dispatchers with detailed real-time information about the current passenger flow and the multi-dimensional impact of waiting decisions in case of train delays. After presenting the algorithmic background and PANDA’s main features, we give a brief online demo. Besides its practical value for train dispatchers, the framework can be used to systematically study scientific questions. Exemplarily, we use our software to experimentally analyse the influence of waiting decisions on realistic passenger flow of Deutsche Bahn. In particular, we evaluate PANDA’s potential benefit for passengers. Our findings indicate that a remarkable reduction in total delay might be possible.

3.25 Mathematical Modelling of Industrialized Timetables

Karl Nachtigall (TU Dresden, DE)

License  Creative Commons BY 3.0 Unported license
© Karl Nachtigall

The computer-added generation of timetables gains noticeably importance to the strategic timetabling in railway transportation due to the application of more efficient mathematical methods and increasing computational capacities. On the one hand the computer added timetabling supports the timetabling specialist through enhancing its work efficiency. On the other hand modern timetabling methods provide the possibility to generate and evaluate several timetables. Transportation and infrastructure companies can thus analyse the available infrastructure with special respect to meaningful traffic-related measures. Therefore, the prompt, automatic and especially conflict-free generation and optimization of timetables is of special significance. The Department of Traffic Flow Science at TU Dresden is developing the programme system TAKT to automatically generate and to mathematically optimize timetables for a predetermined railway infrastructure and an associated operating program to the operator. The solution of the planning problem considers manifold diverse requirements which are partially contrary to one another. Furthermore, TAKT is able to handle large and complex real-world railway networks. This presentation gives an overview of recent research results about the mathematical modelling and solving of industrialized timetables.

3.26 O.R. for conventional rail operations in India

Narayan Rangaraj (Indian Institute of Technology – Mumbai, IN)

License  Creative Commons BY 3.0 Unported license
© Narayan Rangaraj

This talk summarizes O.R. approaches that could be fruitfully tried in the very large conventional rail system in India, especially the passenger transport segment.

In suburban services (in Mumbai, Chennai and Kolkata), there are interesting possibilities in highly resource constrained environments in timetabling and rolling stock planning. The timetables in these places are approximately periodic, but with very many special constraints including platform allocation, that make it quite challenging. Known frameworks arising from PESP may need some additional work to apply in this environment.

In long distance train services, capacity planning using a mix of simulation, scheduling theory and optimization is a potentially useful tool in many congested parts of the network. Passenger trains are handled on sections that carry significant volumes of unscheduled freight traffic (of high revenue potential) and combining goals of punctuality and throughput is quite challenging.

The rail network itself is of very large geographical spread and is very dense in parts. It is not clear what is the best way of managing such a large network in terms of decomposition of zones of control. Area control versus line or route control is one tradeoff that needs to be explored.

In the commercial domain, including in ticketing, quota allocation and revenue management, there has been significant development in the last few years, and there appears to be a lot that the O.R. and Computer Science community can contribute to.

3.27 Train routing selection for the real-time railway traffic management

Marcella Sama (University of Rome III, IT), Andrea D'Ariano, Dario Pacciarelli (University of Rome III, IT), Paola Pellegrini, and Joaquin Rodriguez

License  Creative Commons BY 3.0 Unported license
© Marcella Sama, Andrea D'Ariano, Dario Pacciarelli, Paola Pellegrini, and Joaquin Rodriguez

The growth of demand forces railway infrastructure managers to use the existing infrastructure at full capacity. During daily operations, disturbances may lead to conflicting requests, i.e., time-overlapping requests for the same tracks by multiple trains. The real-time railway traffic management problem (rtRTMP) aims to minimize the delay propagation using simultaneously timing, ordering and routing adjustments, decided in real-time. The problem size and the computational time required to find a good quality solution are strongly affected by the number of alternative routings available to each train. To ease the solution process, we study the train routing selection problem (TRSP). Given a railway network and n trains, the TRSP consists of selecting a subset of alternative routings for each train to be used in the rtRTMP. The TRSP is modelled using a compatibility graph $G = (C, L)$. Each vertex in C associates a train with a feasible routing, each edge in L connects two vertices associated with different trains if the associated routings satisfy possibly existing rolling stock reutilization constraints between the two trains. The compatibility graph is thus n -partite, each component of the partition representing all the alternative routings of a given train. The cost of a vertex represents the potential delay of the associated train when using the related routing, while the cost of an edge estimates the incremental delay due to the resolution of possible conflicts arising when the two trains use the related routings. Solving the TRSP becomes thus the problem of finding m cliques having cardinality n and minimum cost. Preliminary studies of solving the TRSP, using a meta-heuristic approach, have shown very promising impact on the solution of the rtRTMP. Relevant open issues remains to be tackled, such as the definition of helpful properties for solving the TRSP, or the integration between rtRTMP and TRSP.

3.28 Approaches for integrated planning in public transportation

Anita Schöbel (Universität Göttingen, DE)

License  Creative Commons BY 3.0 Unported license
© Anita Schöbel

Approaches for integrated planning in public transportation

Planning of a public transportation system is so far usually done in a sequential way: After the network design, the lines and their frequencies are planned. Based on these, the timetable is set up, and later on the vehicles' schedules and the drivers' schedules. From an optimization point of view such a sequential planning procedure can be regarded as a Greedy approach: in each planning stage one aims at the best one can do. This usually leads to suboptimal solutions. On the other hand, many of these single steps are already NP hard such that solving the integrated problem to optimality seems to be out of scope.

In this talk we review line planning, timetabling and vehicle scheduling and argue that public transportation will benefit from an integrated planning. We sketch ideas and first results on exact algorithms for integrated planning (mainly suitable for special cases).

Moreover, we propose a framework, called “Eigenmodel”, which can be used as basis for designing iterative approaches which re-optimize the line plan, the timetable, or the vehicle schedule while the other input is fixed (see Figure Eigenmodell). We show that such re-optimization procedures lead to new problem instances and illustrate these on re-optimizing line plan and timetable iteratively. In particular, we discuss questions about the convergence of these approaches and end with a list of challenging research questions within the Eigenmodel.

3.29 From Robustness to Resilience – How to evaluate resilience of a timetable

Norio Tomii (*Chiba Institute of Technology, JP*)

License  Creative Commons BY 3.0 Unported license
© Norio Tomii

Recently, resilience of a timetable is attracting considerable attention. Resilience for transportation systems is defined as “the ability to prepare and plan for, absorb, recover from, and more successfully adapt to adverse events.” (Transportation Research Board annual meeting 2015)

From the viewpoint of timetabling, resilience could be interpreted as a timetable which does not give much inconvenience to passengers when a rather big disruption occurs. Thus, we should assume some rescheduling is done, which is a difference from “robustness.” When we assume rescheduling, we have to consider facilities such as track layouts. This means that we have to define resilience for a combination of a timetable (including crew schedules, train-set schedules), a rescheduling algorithm and track layouts. In other words, to achieve resilience of a timetable, we have to integrate a timetable, operation and facilities. There exist a huge amount of such combination and we have to develop an approach to tackle the difficulty of the combinatorial explosion.

We propose an idea to evaluate resilience assuming that a best effort rescheduling is done. We enumerate timetables and facilities and evaluate them using a best-effort rescheduling algorithm. In my presentation, I will introduce some preliminary results for simple cases.

3.30 Decision-support for railway traffic management: What can we actually conclude from previous research work and does the work meet the needs experienced by practitioners?

Johanna Törnquist Krasemann (*Blekinge Institute of Technology – Karlskrona, SE*)

License  Creative Commons BY 3.0 Unported license
© Johanna Törnquist Krasemann

The research efforts dedicated the last 20 years to develop different types of decision support functionalities for railway traffic management is significant. New knowledge and insight on what potential improvements that can be made and what methods there are have been transferred to practitioners. The implementation rate of developed methods is, however, still modest. Interesting questions are therefore, is the focus of the research community addressing the actual needs experienced in practice, and what is needed to bridge between research and practice?

This presentation aims to start a discussion about what we actually can conclude from previous research studies regarding 1) the requirements and needs experienced in practice and 2) the strengths, weaknesses and maturity of existing solution approaches.

3.31 The future of public transport by autonomous vehicles

Lucas Veelenturf (TU Eindhoven, NL) and Afonso Sampaio

License © Creative Commons BY 3.0 Unported license
© Lucas Veelenturf and Afonso Sampaio

The attention on developing autonomous cars is growing rapidly. Multiple high-tech and automotive manufacturers and universities are extensively testing these types of cars. However less attention is paid on how we are going to use this vehicles.

In this talk we assume the autonomous vehicles are available and focus on how to operationalize a public transport with autonomous vehicles. This will make it possible to shift from a supply driven to a demand driven public transport system. Several questions are raised (e.g. How many vehicles are necessary? How to route the vehicles? How to charge the vehicles? etc.)

For the routing problem, we introduced mathematical models based on existing ones in freight transportation like the Dynamic Pick-up and Delivery Problem with Transfers. Both an arc based and a path based formulation were provided of which the path based formulation seemed to be most promising. However, this path based model makes use of a time expanded graph, which can be of an enormous size. Currently we are in the phase of developing fast algorithms to solve this problem via the path based model.

3.32 Engineering Graph-Based Models for Dynamic Timetable Information Systems

Christos Zaroliagis (CTI & University of Patras, GR)

License © Creative Commons BY 3.0 Unported license
© Christos Zaroliagis

Joint work of Alessio Cionini, Gianlorenzo D'Angelo, Mattia D'Emidio, Daniele Frigioni, Kalliopi Giannakopoulou, and Andreas Paraskevopoulos

Main reference A. Cionini, G. D'Angelo, M. D'Emidio, D. Frigioni, K. Giannakopoulou, A. Paraskevopoulos, C. D. Zaroliagis, "Engineering Graph-Based Models for Dynamic Timetable Information Systems", in Proc. of the 14th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS'14), OASICS, Vol. 42, pp. 46–61, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014.

URL <http://dx.doi.org/10.4230/OASICS.ATMOS.2014.46>

Many efforts have been done in the last years to develop efficient algorithmic solutions for the problem of computing *best routes* in schedule-based public transportation systems. Advances in this area have been remarkable: nowadays, we have models to represent the given input timetable that allow us to answer queries for optimal journeys in few milliseconds, also at a very large scale. Such models can be broadly classified into two types: those representing the timetable as an array, and those representing it as a graph. Array-based models have been shown to be very effective in terms of query time, while graph-based ones usually answer queries by computing shortest paths, and hence they are suitable to be combined with the (very effective) speed-up techniques developed for road networks in the recent past.

In this paper, we study the behavior of graph-based models in the prominent case of dynamic scenarios, i.e., when delays might (unpredictably) occur to the original timetable. In particular, we make the following contributions. First, we consider the graph-based reduced time-expanded model and give a simplified and optimized routine for handling delays, and a re-engineered and fine-tuned query algorithm. Second, we propose a new graph-based model, namely the Dynamic Timetable Model, natively tailored to efficiently incorporate dynamic updates, along with a query algorithm and a routine for handling delays. Third, we show how to adapt the unidirectional ALT algorithm to such graph-based models. We have chosen this speed-up technique since it supports dynamic changes, and a careful implementation of it can significantly boost its performance. Finally, we provide an experimental study to assess the effectiveness of all proposed models and algorithms and to compare them with the state of the art. We evaluate both new and existing approaches by implementing and testing them on real-world timetables subject to synthetic delays.

Our experimental results show that: (i) the Dynamic Timetable Model is the best model in terms of computational time required for handling delays; (ii) graph-based models are competitive to array-based models with respect to query time; (iii) the Dynamic Timetable Model compares favorably with both the original and the reduced time-expanded model regarding space; (iv) combining the graph-based models with some speed-up techniques designed for road networks, such as ALT, is a very promising approach.

Participants

- Ralf Borndörfer
Konrad-Zuse-Zentrum –
Berlin, DE
- Valentina Cacchiani
University of Bologna, IT
- Francesco Corman
TU Delft, NL
- David De Almeida
SNCF – Paris, FR
- Markus Friedrich
Universität Stuttgart, DE
- Marc Goerigk
Lancaster University, GB
- Rob Goverde
TU Delft, NL
- Jonas Harbering
Universität Göttingen, DE
- Mark Hickman
The Univ. of Queensland, AU
- Dennis Huisman
Erasmus Univ. – Rotterdam, NL
- Giuseppe F. Italiano
University of Rome “Tor
Vergata”, IT
- Natalia Kliewer
FU Berlin, DE
- Leo G. Kroon
Erasmus Univ. – Rotterdam, NL
- Allan Larsen
Technical Univ. of Denmark –
Lyngby, DK
- Jesper Larsen
Technical Univ. of Denmark –
Lyngby, DK
- Marco Laumanns
IBM Research Zurich, CH
- Janny Leung
The Chinese University of Hong
Kong, HK
- Marco Lübbecke
RWTH Aachen, DE
- Carlo Mannino
SINTEF ICT – Oslo, NO
- Gabor Maroti
VU University Amsterdam, NL
- Lingyun Meng
Beijing Jiaotong University, CN
- Juan Antonio Mesa
University of Sevilla, ES
- Matus Mihalak
Maastricht University, NL
- Pitu Mirchandani
Arizona State University –
Tempe, US
- Rolf H. Möhring
TU Berlin, DE
- Matthias Müller-Hannemann
Martin-Luther-Universität
Halle-Wittenberg, DE
- Karl Nachtigall
TU Dresden, DE
- Morten Nyhave Nielsen
DSB – Copenhagen, DK
- Dario Pacciarelli
University of Rome III, IT
- Thomas Pajor
Cupertino, US
- Narayan Rangaraj
Indian Institute of Technology –
Mumbai, IN
- Marcella Sama
University of Rome III, IT
- Anita Schöbel
Universität Göttingen, DE
- Leena Suhl
Universität Paderborn, DE
- Johanna Törnquist Krasemann
Blekinge Institute of Technology –
Karlskrona, SE
- Norio Tomii
Chiba Institute of Technology, JP
- Lucas Veelenturf
TU Eindhoven, NL
- Dorothea Wagner
KIT – Karlsruher Institut für
Technologie, DE
- Christos Zaroliagis
CTI & University of Patras, GR



Machine Learning for Dynamic Software Analysis: Potentials and Limits

Edited by

**Amel Bennaceur¹, Dimitra Giannakopoulou², Reiner Hähnle³, and
Karl Meinke⁴**

1 The Open University – Milton Keynes, GB, amel.bennaceur@open.ac.uk

2 NASA – Moffett Field, US, dimitra.Giannakopoulou@nasa.gov

3 TU Darmstadt, DE, haehnle@cs.tu-darmstadt.de

4 KTH Royal Institute of Technology – Stockholm, SE, karlm@kth.se

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16172 “Machine Learning for Dynamic Software Analysis: Potentials and Limits”. Machine learning is a powerful paradigm for software analysis that provides novel approaches to automating the generation of models and other essential artefacts. This Dagstuhl Seminar brought together top researchers active in the fields of machine learning and software analysis to have a better understanding of the synergies between these fields and suggest new directions and collaborations for future research.

Seminar April 24–27, 2016 – <http://www.dagstuhl.de/16172>

1998 ACM Subject Classification D.2 Software Engineering, D.2.4 Software/Program Verification, I.2 Artificial Intelligence, I.2.6 Learning

Keywords and phrases Machine learning, Automata learning, Software analysis, Dynamic analysis, Testing, Model extraction, Systems integration

Digital Object Identifier 10.4230/DagRep.6.4.161

1 Executive Summary

Amel Bennaceur

Reiner Hähnle

Karl Meinke

License © Creative Commons BY 3.0 Unported license
© Amel Bennaceur, Reiner Hähnle, and Karl Meinke

Machine learning of software artefacts is an emerging area of interaction between the machine learning (ML) and software analysis (SA) communities. Increased productivity in software engineering hinges on the creation of new adaptive, scalable tools that can analyse large and continuously changing software systems. For example: agile software development using continuous integration and delivery can require new documentation models, static analyses, proofs and tests of millions of lines of code every 24 hours. These needs are being addressed by new SA techniques based on machine learning, such as learning-based software testing, invariant generation or code synthesis.

Machine learning is a powerful paradigm for SA that provides novel approaches to automating the generation of models and other essential artefacts. However, the ML and SA communities are traditionally separate, each with its own agenda. This Dagstuhl Seminar



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

Machine Learning for Dynamic Software Analysis: Potentials and Limits, *Dagstuhl Reports*, Vol. 6, Issue 4, pp. 161–173

Editors: Amel Bennaceur, Dimitra Giannakopoulou, Reiner Hähnle, and Karl Meinke



DAGSTUHL
REPORTS Dagstuhl Reports

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

brought together top researchers active in these two fields who can present the state of the art, and suggest new directions and collaborations for future research. We, the organisers, feel strongly that both communities have much to learn from each other, and the seminar focused strongly on fostering a spirit of collaboration.

The first day was dedicated to mutual education through a series of tutorials by leading researchers in both ML and SA to familiarise everyone with the terminology, research methodologies, and main approach of each community. The second day was dedicated to brainstorming and focused discussion in small groups, each of which supported by one of the organisers acting as a facilitator. At the end of the day a plenary session was held for each group to share a summary of their discussions. The participants also reflected and compared their findings. The morning of the third day was dedicated to the integration of the groups and further planning.

This report presents an overview of the talks given at the seminar and summaries of the discussions of the participants.

Acknowledgements. The organisers would like to express their gratitude to the participants and the Schloss Dagstuhl team for a productive and exciting seminar.

2 Table of Contents

Executive Summary

Amel Bennaceur, Reiner Hähnle, and Karl Meinke 161

Overview of Talks

Machine Learning for Emergent Middleware <i>Amel Bennaceur</i>	164
Learning Register Automata Models <i>Falk Howar</i>	164
Static (Software) Analysis <i>Reiner Hähnle</i>	164
Learning-based Testing: Recent Progress and Future Prospects <i>Karl Meinke</i>	165
Towards Automata Learning in Practice <i>Bernhard Steffen</i>	165
Learning State Machines <i>Sicco Verwer</i>	166

Working groups

Different Kinds of Models <i>Andreas Abel, Amel Bennaceur, Roland Groz, Falk Howar, Frits Vaandrager, Sicco Verwer, and Neil Walkinshaw</i>	166
Combining White-Box and Glass-Box Analysis <i>Falk Howar, Andreas Abel, Pavol Bielik, Radu Grosu, Roland Groz, Reiner Hähnle, Bengt Jonsson, Mohammad Reza Mousavi, Zvonimir Rakamaric, Alessandra Russo, Sicco Verwer, and Andrzej Wasowski</i>	167
Machine learning for System Composition <i>Falk Howar, Amel Bennaceur, Bengt Jonsson, Alessandra Russo, Sicco Verwer, and Andrzej Wasowski</i>	168
Combination of Static Analysis and Learning <i>Bernhard Steffen, Pavol Bielik, Radu Grosu, Reiner Hähnle, Bengt Jonsson, Mohammad Reza Mousavi, Daniel Neider, and Zvonimir Rakamaric</i>	169
Benchmark Building and Sharing <i>Frits Vaandrager, Dalal Alrajeh, Amel Bennaceur, Roland Groz, Karl Meinke, Daniel Neider, Bernhard Steffen, and Neil Walkinshaw</i>	170
Learning and Testing <i>Neil Walkinshaw, Andreas Abel, Dalal Alrajeh, Pavol Bielik, Roland Groz, Reiner Hähnle, Karl Meinke, Mohammad Reza Mousavi, Daniel Neider, Zvonimir Rakamaric, Bernhard Steffen, and Frits Vaandrager</i>	171

Participants 173

3 Overview of Talks

3.1 Machine Learning for Emergent Middleware

Amel Bennaceur (The Open University – Milton Keynes, GB)

License  Creative Commons BY 3.0 Unported license
© Amel Bennaceur

Highly dynamic and heterogeneous distributed systems are challenging today’s middleware technologies. Existing middleware paradigms are unable to deliver on their most central promise, which is offering interoperability. In this talk, I argue for the need to dynamically synthesise distributed system infrastructures according to the current operating environment, thereby generating “Emergent Middleware” to mediate interactions among heterogeneous networked systems that interact in an ad hoc way. I will explain the overall architecture underlying Emergent Middleware, and in particular focuses on the key role of learning in supporting such a process, spanning statistical learning to infer the semantics of networked system functions and automata learning to extract the related behaviours of networked systems.

3.2 Learning Register Automata Models

Falk Howar (IPSSE – Goslar, DE)

License  Creative Commons BY 3.0 Unported license
© Falk Howar

Learning algorithms for register automata infer models with parameterized actions, symbolic guards, and memory. In this talk, we give a brief overview of how active automata learning has been extended over the past decade to infer such richer models. We focus on one line of work that is based on a generalized Myhill-Nerode theorem for data languages and the inference of symbolic decision trees from test cases. We present some key insights and open questions from this line of work and compare it to other approaches that tackle the same problem.

3.3 Static (Software) Analysis

Reiner Hähnle (TU Darmstadt, DE)

License  Creative Commons BY 3.0 Unported license
© Reiner Hähnle

We give a brief introduction into the field of static analysis: definition, scope, techniques, challenges, trends. We also juxtapose static analysis of software with dynamic based on learning. We compare their relative strengths and weaknesses and try to work out where the opportunities for their possible combination lie.

3.4 Learning-based Testing: Recent Progress and Future Prospects

Karl Meinke (KTH Royal Institute of Technology – Stockholm, SE)

License  Creative Commons BY 3.0 Unported license
© Karl Meinke

We present a survey of recent progress in the area of learning-based testing (LBT). The emphasis is primarily on fundamental concepts and theoretical principles, rather than applications and case studies. After surveying the basic principles and a concrete implementation of the approach, we describe recent directions in research such as: quantifying the hardness of learning problems, over-approximation methods for learning, and quantifying the power of model checker generated test cases. The common theme underlying these research directions is seen to be metrics for model convergence. Such metrics enable a precise, general and quantitative approach to both speed of learning and test coverage. Moreover, quantitative approaches to black-box test coverage serve to distinguish LBT from alternative approaches such as random and search-based testing. We conclude by outlining some prospects for future research.

3.5 Towards Automata Learning in Practice

Bernhard Steffen (TU Dortmund, DE)

License  Creative Commons BY 3.0 Unported license
© Bernhard Steffen

In the last decade automata learning has attracted practical attention in software engineering e.g. as a means to lower the hurdle of so-called model-based testing, by overcoming the problem of the required a priori models, or as a way to mine runnable (legacy) software for behavioural specifications. In order to achieve true practicality, automata learning has 1) to increase scalability, 2) to move from its original theoretical roots, which focused on standard finite state machines to more expressive formalisms, and finally 3) to establish notions of quality assurance. Concerning 1) and 3), the TTT algorithm [1, 2] is promising, as it provides a scalable solution for dealing with extremely long counter examples, which are characteristic for the so-called life-long learning paradigm. Rather than making explicit quality statements, this paradigm establishes a continuous improvement cycle, and therefore an approach to quality assurance adequate, in particular, for agile software development. 2) has been addressed e.g. via extensions to Register Automata [3] and extended finite State Machines [4] (cf. the contributions of Falk Howar to the seminar). The Open-Sources Learnlib [5] aims at making all these algorithms available to the public.

References

- 1 Malte Isberner, Falk Howar, Bernhard Steffen: The TTT Algorithm: A Redundancy-Free Approach to Active Automata Learning. *RV* 2014:307–322
- 2 Malte Isberner Foundations of active Automata Learning: an algorithmic perspective (PhD thesis, Dortmund, 2015)
- 3 Malte Isberner, Falk Howar, Bernhard Steffen: Learning register automata: from languages to program structures. *Machine Learning* 96(1–2):65–98 (2014)
- 4 Sofia Cassel, Falk Howar, Bengt Jonsson, Bernhard Steffen: Active learning for extended finite state machines. *Formal Asp. Comput.* 28(2):233–263 (2016)
- 5 Malte Isberner, Falk Howar, Bernhard Steffen: The Open-Source LearnLib – A Framework for Active Automata Learning. *CAV* (1) 2015:487–495

3.6 Learning State Machines

Sicco Verwer (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Sicco Verwer

This tutorial is dedicated to learning complex state machines (including timing and parameters), using learning for more than just prediction (for instance model checking), and the power of search methods (using SAT-solvers and Mixed Integer Programming) in machine learning.

4 Working groups

4.1 Different Kinds of Models

Andreas Abel (Universität des Saarlandes – Saarbrücken, DE), Amel Bennaceur (The Open University – Milton Keynes, GB), Roland Groz (LIG – Grenoble, FR), Falk Howar (IPSSE – Goslar, DE), Frits Vaandrager (Radboud University Nijmegen, NL), Sicco Verwer (TU Delft, NL), and Neil Walkinshaw (University of Leicester, GB)

License  Creative Commons BY 3.0 Unported license
© Andreas Abel, Amel Bennaceur, Roland Groz, Falk Howar, Frits Vaandrager, Sicco Verwer, and Neil Walkinshaw

In this break-out session we discussed which models should we target for learning: what kind of models should we learn, and in particular what are the challenges for learning extended finite state models.

Since all participants in the group were dealing with state based models, we mostly discussed the issues related to this type of model, although we acknowledged that logic based models, in particular rule-based models, are just as worthy of interest for software engineering.

The type of mode may depend on the context in which the learning is done, the assumptions on what is available for learning (do we have negative as well as positive samples, do we have an oracle, can we reset or checkpoint a system ?) as well as the intended use of the models: is the model meant for testing, for static analysis, for reverse engineering or documenting etc.

Regarding the flavours of state models that can be learnt, we reckoned that many different kinds of models are already supported: Mealy machines and DFA are the most common, but we also have Moore machines, register automata, combination with rule based systems, hybrid automata. We could also consider LTS, IOTS, timed models. It is also interesting in some contexts to learn non-deterministic models of deterministic systems, and similarly to learn stochastic models. The main frontier is about what kind of extended state machine models could we learn, more powerful than register automata. Another so far not addressed issue is that of concurrency models, for which e.g. Petri Nets could be a target.

The major challenge for model learning in software analysis is about learning parametrised actions and associated data relations. We reckon that register automata are still too limited, we need richer models. At the same time, it is important to be able to identify relevant abstractions from data. Typically, being able to extract a parametrised model of actions with the relevant parameters from a traffic capture, without the pains of having to have a human expert doing it and writing the corresponding adaptor (aka mapper) is a key challenge. One direction can be to use statistical learning methods, such as PCA (Principal Component

Analysis). In this context, it is also important to consider methods that can learn in the presence of noisy data.

Other challenges have also been discussed: it is important to learn not just a system, but also a model of its environment. For a wide use of model learning, it is also important to be able to learn models that address non-functional characteristics of software such as performance or security. For a large number of applications, it is also important to learn understandable models, that can easily be interpreted. In this view, it may be better to learn simpler, more abstract models than more accurate ones that could be too complex.

Throughout the seminar, we considered a crucial issue: can learning end up with a consistent approximation: probably an over approximation, or maybe an under approximation? Most state based learning approaches produce models that are neither an over- nor an under-approximation. Rule based systems are more sensitive to the notion of monotonicity. But maybe the issue of having a consistent over-approximation is only useful in the context of verification, where we need a boolean answer. Another direction is to consider accuracy, as in the PAC concept (Probably Approximately Correct).

Finally, other issues with models have also been discussed. First, with most current tools, handling of timing issues, esp. events caused by a timeout, is done in ad hoc manner. Research is still need to adapt learning with a sound model for the passage of time. We also discussed the use of causality for identifying data relations. The idea is to use variations on some input parameters to check their influence on output parameters.

4.2 Combining White-Box and Glass-Box Analysis

Falk Howar (IPSSSE – Goslar, DE), Andreas Abel (Universität des Saarlandes – Saarbrücken, DE), Pavol Bielik (ETH Zürich, CH), Radu Grosu (TU Wien, AT), Roland Groz (LIG – Grenoble, FR), Reiner Hähnle (TU Darmstadt, DE), Bengt Jonsson (Uppsala University, SE), Mohammad Reza Mousavi (Halmstad University, SE), Zvonimir Rakamaric (University of Utah – Salt Lake City, US), Alessandra Russo (Imperial College London, GB), Sicco Verwer (TU Delft, NL), and Andrzej Wasowski (IT University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license

© Falk Howar, Andreas Abel, Pavol Bielik, Radu Grosu, Roland Groz, Reiner Hähnle, Bengt Jonsson, Mohammad Reza Mousavi, Zvonimir Rakamaric, Alessandra Russo, Sicco Verwer, and Andrzej Wasowski

In this session, we explored ideas for combining black-box and analyses glass-box analyses in a meaningful way that would benefit either one analysis technique, or lead to a new and more powerful combined approach. The session identified three potential scenarios for the integration of approaches from both worlds, and discussed resulting practical challenges.

Learning for Environment Generation

The first idea was using learned models for verification: a model of a component's environment could be used to close the (open) component for verification. The main challenge in this scenario is providing quality information or correctness guarantees for models. Without such guarantees verification may not be sound, but it may still be able to find bugs.

Using Glass-Box Techniques in Learning

A second idea was using glass-box techniques in automata learning. There exist already some works that incorporate domain information generated by static analysis (e.g., for partial order reduction on the alphabet). Other glass-box techniques may be useful for as a basis for implementing equivalence queries or new kinds of queries that can speed up the learning process (e.g., that after some prefix a certain behaviour is never observable).

Integrating Learning for Glass-Box Analysis

The third idea was a tight integration of glass-box and black-box techniques: Static analysis could profit from dynamic analysis when it fails to produce good enough results. At the same time, the results produced by a learning algorithm could be improved by using static analysis to determine what to expose of a system during learning. The main challenge in this scenario is defining an interface for exchanging information between the two analysis techniques.

4.3 Machine learning for System Composition

Falk Howar (IPSSE – Goslar, DE), Amel Bennaceur (The Open University – Milton Keynes, GB), Bengt Jonsson (Uppsala University, SE), Alessandra Russo (Imperial College London, GB), Sicco Verwer (TU Delft, NL), and Andrzej Wasowski (IT University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license
© Falk Howar, Amel Bennaceur, Bengt Jonsson, Alessandra Russo, Sicco Verwer, and Andrzej Wasowski

The topic of machine learning for system composition was discussed in this break-out session. Machine Learning approach, in this context, is primarily intended to be automata learning. The system composition problem is therefore how to combine together automata that describe behaviours of components of a system. Two possible approaches were identified. On one hand, the union of the languages of the different components can be considered and the automata of the entire system can be computed using automata learning mechanisms. But clearly this may have scalability problems. A second approach would be to start from individual automata of each component and then learn mediator models that allow the composition of the individual automata. Learning automata can in this case be used for both learning individual automata as well as learning mediator's automata. On the other hand, the area of learning state machine has also seen work in the context of logic-based learning, where the objectives is to learn temporal specification that can be translated into labelled transition systems that cover given positive traces and do not cover negative traces. Interesting question is there whether there is a potential synergy between logic-based learning and learning automata that could lead to novel mechanisms and/or improvements of state machine learning processes. Main features of logic-based learning include knowledge about the problem in hand. For instance they can include general knowledge about state merge; they take in input labelled positive and negative traces; and they are capable of learning linear temporal logic descriptions that, when translated into labelled transition system behaviour models, are guaranteed to accept the positive traces and reject all the negative traces. Finally they learn within the search space of a defined language bias, which can declaratively constraints with domain-specific expert knowledge. So, one of the open questions is how the two approaches relate with each

other. Are there similarities between learning automata and logic-based learning that can be exploited to allow synergisms between the two types of ML approaches in the context of software analysis. It has been pointed out that earlier work exists on how to translate State Machine language into first-order (FO) logic knowledge representation language IDP. This FO formula gets grounded to a propositional satisfiability, which is solved using a SAT solver. The LBL approach is similar but aims to find an FO formula directly, which gets translated into a labeled transition system after learning takes place. So, a very promising direction for research is to combine these methods, adding the capability of including expert-knowledge as constraints to the state machine learning process. A synergy between logic-based learning and learning automata may also lead to the development of distributed learning algorithms that allow models of system components to be learned collaboratively in a manner that guarantee properties of the composition expressed as constraints in the expert knowledge of the learner of each component. For the learning of mediator models, it would be interesting to try to use logic-based learning to infer the mediator (glue code) between existing components. In this case, the descriptions of the components represent a model of the environment while the desirable properties of the interactions represents the goal model from which a declarative specification that gives the mediator automata could be learned. We could then compare/contrast the result with that obtained by using automata learning.

4.4 Combination of Static Analysis and Learning

Bernhard Steffen (TU Dortmund, DE), Pavol Bielik (ETH Zürich, CH), Radu Grosu (TU Wien, AT), Reiner Hähnle (TU Darmstadt, DE), Bengt Jonsson (Uppsala University, SE), Mohammad Reza Mousavi (Halmstad University, SE), Daniel Neider (Los Angeles, US), and Zvonimir Rakamaric (University of Utah – Salt Lake City, US)

License © Creative Commons BY 3.0 Unported license

© Bernhard Steffen, Pavol Bielik, Radu Grosu, Reiner Hähnle, Bengt Jonsson, Mohammad Reza Mousavi, Daniel Neider, and Zvonimir Rakamaric

In particular, because of the previous discussion in the Discussion Group on Black Box and White Box Methods, the discussion here focused on two topics:

1. How to combine concolic execution and learning. As a success [1] was mentioned which uses the (concrete) access sequences of a learned hypothesis as a means to provide concrete values to drive the concolic/symbolic execution by inserting fitting concrete values into the symbolic execution process. Essentially this means that one uses automata learning as a supportive oracle for providing adequate concrete values for concolic execution. On the other side, the concolic or symbolic execution may support the test-based search for counter examples (the typical practical realization of the so-called equivalence queries) by reducing the search space. A more general underlying question here is how to synchronize the two worlds, meaning how symbolic states that arise during symbolic execution can be related to the states of a learned hypothesis. This relation hinges on a common notational understanding in terms of an adequate abstraction level. Such a level may be revealed via the search for e.g. key variables (e.g. those that strongly influence the control flow) using ‘classical’ static analysis techniques (slicing etc.).
2. In contrast to static analysis and typical other white box techniques, learning does not provide any guarantees like over approximation or under approximation. We discussed the question of whether it might be possible use static analysis to guarantee over-approximation during the learning process. It seems that reducing the chain of hypothesis

to only contain over approximations almost inevitably breaks termination in general. Some compare automata learning with polynomial interpolation, as it like automata learning:

- Provides precise answers if the function to be approximated is indeed a polynomial of degree n , the degree of the current interpolation (just replace degree by number of states).
- There is in general no way to guarantee that the polynomial is an over or upper approximation of the target function.

The way proposed during the discussion to replace all the parts that are not really known by ‘chaos’ (i.e. the process which can do whatever it likes), reminds me of setting all function value outside the supported set to infinity, which certainly guarantees over approximation, but, in a way, destroys the charm of the interpolation idea, which consists of leveraging finite information to obtain something infinite (unfortunately at the cost of the abovementioned guarantees). However, like in polynomial interpolation, one may think of metrics and ways of error approximation (here in terms of probabilities). It still has to be seen, how practical such approaches might be.

An alternative approach could be to not guarantee over approximation continuously, but only on demand, and this approach could well profit from static analysis, showing that all the abstractly feasible paths are indeed covered by the hypothesis.

References

- 1 CY Cho, D Babic, P Poosankam, KZ Chen, EXJ Wu, D Song *MACE: Model-inference-Assisted Concolic Exploration for Protocol and Vulnerability Discovery*. USENIX Security Symposium, pp. 139–154. 2011

4.5 Benchmark Building and Sharing

Frits Vaandrager (Radboud University Nijmegen, NL), Dalal Alrajeh (Imperial College London, GB), Amel Bennaceur (The Open University – Milton Keynes, GB), Roland Groz (LIG – Grenoble, FR), Karl Meinke (KTH Royal Institute of Technology – Stockholm, SE), Daniel Neider (Los Angeles, US), Bernhard Steffen (TU Dortmund, DE), and Neil Walkinshaw (University of Leicester, GB)

License © Creative Commons BY 3.0 Unported license
 © Frits Vaandrager, Dalal Alrajeh, Amel Bennaceur, Roland Groz, Karl Meinke, Daniel Neider, Bernhard Steffen, and Neil Walkinshaw

We had an interesting discussion on the use of benchmarks for automata learning and testing. Radboud University has set up a repository for Mealy machines and register automata. Frits Vaandrager encouraged everybody to submit benchmarks. Extension to other classes of models such as Moore machines, DFAs,.. are encouraged. All participants agreed that benchmarks are important and useful. Benchmarks measure the relative performance of different approaches/systems and they are important to check whether tools and methods advance and whether new methods are effective. Systematic use of benchmarks is a sign of maturity of a scientific field. Several participants offered to contribute benchmarks to the Radboud University repository. The discussion made it clear that (a) we have different types of benchmarks, and (b) many different criteria for evaluation of algorithms and tools.

Different types of benchmarks:

- Benchmarks for evaluating efficiency of algorithms and tools
- Challenges for pushing the state-of-the-art (e.g. RERS challenges)
- Benchmarks for illustrating usefulness of a method or tool.

Criteria for evaluation of algorithms and tools:

- Does the tool aim at fully learning the benchmark or just at giving suitable aggregate information of data that have been gathered. E.g. there are tools for invariant generation.
- Number of inputs events
- Number of test sequences
- Wall clock time needed for learning (reset or certain inputs may require lot of time)
- Quality of intermediate hypothesis; how long it takes before you get first reasonable model
- How interpretable are the results (e.g. by discovering hierarchy and parallel composition)
- How easy it is to parallelize learning

We agreed to elaborate this list; when people come with new tools/algorithms they should make it clear, using the benchmarks, at which points they are improving the state-of-the-art. We also discussed formats for the benchmarks: For Mealy machines dot files appear to be ok. For register automata, Fides Aarts and Falk Howar have proposed a format. Since in the area of EFSMs things have not stabilized yet standardization may be a mature field is characterized by the presence of two other types of evaluation: Substantial (industrial) case studies are important to check whether tools/methods are relevant for practical problems and do sufficiently scale. Experimental usability studies are needed to find out whether a new method/tool can be integrated into practical workflow with advantage.

4.6 Learning and Testing

Neil Walkinshaw (University of Leicester, GB), Andreas Abel (Universität des Saarlandes – Saarbrücken, DE), Dalal Alrajeh (Imperial College London, GB), Pavol Bielik (ETH Zürich, CH), Roland Groz (LIG – Grenoble, FR), Reiner Hähnle (TU Darmstadt, DE), Karl Meinke (KTH Royal Institute of Technology – Stockholm, SE), Mohammad Reza Mousavi (Halmstad University, SE), Daniel Neider (Los Angeles, US), Zvonimir Rakamaric (University of Utah – Salt Lake City, US), Bernhard Steffen (TU Dortmund, DE), and Frits Vaandrager (Radboud University Nijmegen, NL)

License © Creative Commons BY 3.0 Unported license

© Neil Walkinshaw, Andreas Abel, Dalal Alrajeh, Pavol Bielik, Roland Groz, Reiner Hähnle, Karl Meinke, Mohammad Reza Mousavi, Daniel Neider, Zvonimir Rakamaric, Bernhard Steffen, and Frits Vaandrager

We had an active discussion on the combination of testing and inference. The discussion was structured according to SWOT (Strengths, Weaknesses, Opportunities and Threats), to provide an idea of the state-of-the-art, and where things could go from here. These are summarised individually in the following paragraphs.

Strengths: Automation is a key strength, addressing the key weakness with conventional Model Based Testing (the need to invest effort into producing a model). The automation is especially pronounced when the models that are inferred are associated with established testing algorithms (as is the case for Finite State Machines). Ultimately, the approach does require a degree of manual intervention (e.g. to validate test outputs), but the point was made that another strength of the approach is that this actually dovetails nicely with iterative, agile techniques. If the underlying development process is iterative, inference-based testing can be used to automatically generate test inputs, whilst the validation can occur in its normal setting. There was also the observation that current approaches that combine testing

and inference tend to extract value from established test sets – the existing tests provide the basis for the (initial) model inference.

Weaknesses: Current efforts to combine model inference and testing still suffer from numerous weaknesses, presenting a fertile basis for future research. They do not tend to scale well, often requiring an unrealistic number of test executions or other forms of user input. The currently inferred models tend to have limited expressivity (being predominantly FSMs). In settings where there are different potential Machine Learners, the task of choosing a given learner is often still a matter of intuition. Then there is also the fact that current efforts at empirical evaluation are often very limited; they are rarely compared against random testing, rarely involve genuine / seeded faults, often focus on small-scale systems, and are often based upon questionable metrics of accuracy and efficiency. There is also the broader question of what it means when a test-run has “finished” – there has been little discussion of how much assurance can be derived from this. Finally, there is the task of implementation; the challenge of abridging what are often complex Machine Learning systems with a fully fledged testing engine is challenging, and tools can accordingly rarely be easily deployed.

Opportunities: There is an extensive interest from industry; the problems addressed by the combination of testing and inference are timely. There are also several avenues by which additional knowledge can be embedded into the process (e.g. domain knowledge to prevent inference mistakes), which could easily address some of the aforementioned weaknesses. Product-lines offer an interesting, more controlled environment within which to develop test/inference systems. If the approaches are used as the basis for regression testing, this can remove a lot of the manual effort required for validating test outputs (because this can be checked against previous versions instead).

Threats: Happily, the room could only think of few threats. The top one was that, although industry is generally enthusiastic about the idea of automated testing and the opportunities that ML brings, there are often unrealistic expectations. This can lead to frustration, e.g. when it comes to the realisation that abstractions have to be generated.

Participants

- Andreas Abel
Universität des Saarlandes –
Saarbrücken, DE
- Dalal Alrajeh
Imperial College London, GB
- Amel Bennaceur
The Open University - Milton
Keynes, GB
- Pavol Bielik
ETH Zürich, CH
- Radu Grosu
TU Wien, AT
- Roland Groz
LIG – Grenoble, FR
- Reiner Hähnle
TU Darmstadt, DE
- Falk Howar
IPSSE – Goslar, DE
- Bengt Jonsson
Uppsala University, SE
- Karl Meinke
KTH Royal Institute of
Technology – Stockholm, SE
- Mohammad Reza Mousavi
Halmstad University, SE
- Daniel Neider
Los Angeles, US
- Zvonimir Rakamaric
University of Utah – Salt Lake
City, US
- Alessandra Russo
Imperial College London, GB
- Bernhard Steffen
TU Dortmund, DE
- Frits Vaandrager
Radboud Univ. Nijmegen, NL
- Sicco Verwer
TU Delft, NL
- Neil Walkinshaw
University of Leicester, GB
- Andrzej Wasowski
IT Univ. of Copenhagen, DK

