



DAGSTUHL REPORTS

Volume 6, Issue 10, October 2016

| | |
|--|-----|
| Programming Language Techniques for Incremental and Reactive Computing (Dagstuhl Seminar 16402) <i>Camil Demetrescu, Sebastian Erdweg, Matthew A. Hammer, and Shriram Krishnamurthi</i> | 1 |
| Algebraic Methods in Computational Complexity (Dagstuhl Seminar 16411) <i>Valentine Kabanets, Thomas Thierauf, Jacobo Torán, and Christopher Umans</i> | 13 |
| Automated Algorithm Selection and Configuration (Dagstuhl Seminar 16412) <i>Holger H. Hoos, Frank Neumann, and Heike Trautmann</i> | 33 |
| Universality of Proofs (Dagstuhl Seminar 16421) <i>Gilles Dowek, Catherine Dubois, Brigitte Pientka, and Florian Rabe</i> | 75 |
| Computation over Compressed Structured Data (Dagstuhl Seminar 16431) <i>Philip Bille, Markus Lohrey, Sebastian Maneth, and Gonzalo Navarro</i> | 99 |
| Adaptive Isolation for Predictability and Security (Dagstuhl Seminar 16441) <i>Tulika Mitra, Jürgen Teich, and Lothar Thiele</i> | 120 |
| Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR) (Dagstuhl Seminar 16442) <i>Roger K. Moore, Serge Thill, and Ricard Marxer</i> | 154 |

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

March, 2017

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.6.10.i

Programming Language Techniques for Incremental and Reactive Computing

Edited by

**Camil Demetrescu¹, Sebastian Erdweg², Matthew A. Hammer³,
and Shriram Krishnamurthi⁴**

1 Sapienza University of Rome, IT, demetres@dis.uniroma1.it

2 TU Delft, NL, s.t.erdweg@tudelft.nl

3 University of Colorado – Boulder, US, matthew.hammer@colorado.edu

4 Brown University – Providence, US, sk@cs.brown.edu

Abstract

Incremental computations are those that process input changes faster than naive computation that runs from scratch, and reactive computations consist of interactive behavior that varies over time. Due to the importance and prevalence of incremental, reactive systems, ad hoc variants of incremental and reactive computation are ubiquitous in modern software systems.

In response to this reality, the PL research community has worked for several decades to advance new languages for systems that interface with a dynamically-changing environment. In this space, researchers propose new general-purpose languages and algorithms to express and implement efficient, dynamic behavior, in the form of incremental and reactive language systems.

While these research lines continue to develop successfully, this work lacks a shared community that synthesizes a collective discussion about common motivations, alternative techniques, current results and future challenges. To overcome this lack of community, this seminar will work towards building one, by strengthening existing research connections and by forging new ones. Developing a shared culture is critical to the future advancement of incremental and reactive computing in modern PL research, and in turn, this PL research is critical to developing the efficient, understandable interactive systems of the future.

Seminar October 3–7, 2016 – <http://www.dagstuhl.de/16402>

1998 ACM Subject Classification F.3.2 Semantics of Programming Languages, F.3.3 Studies of Program Constructs, F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Incremental computing, reactive programming, memoization, change propagation, dynamic dependency graph, dataflow programming, live programming

Digital Object Identifier 10.4230/DagRep.6.10.1

1 Executive Summary

Matthew A. Hammer

License  Creative Commons BY 3.0 Unported license
© Matthew A. Hammer

We sought to hold a Dagstuhl Seminar that would bring together programming language (PL) researchers focusing on incremental and reactive computing behavior. The meta-level purpose of this seminar was to take an initial step toward developing a community of experts from the disparate threads of successful research. In that this seminar provoked discussion about common and differing motivations, techniques, and future challenges, this event was successful in starting to cultivate this culture.



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

Programming Language Techniques for Incremental and Reactive Computing, *Dagstuhl Reports*, Vol. 6, Issue 10, pp. 1–12

Editors: Camil Demetrescu, Sebastian Erdweg, Matthew A. Hammer, and Shriram Krishnamurthi



DAGSTUHL
REPORTS Dagstuhl Reports

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

Short-term concrete outcomes: Thus far, there are been two concrete outcomes of this seminar:

1. *Wikipedia article outlines and edits* (Section 3.3)
2. *First Workshop on Incremental Computation (IC) at PLDI 2017* (Section 5)

Section 3 gives an overview of the event structure of the seminar, and details some of the event's outcomes, including outline brainstorming and Wikipedia editing, and the creation of a new Workshop on Incremental Computing (IC). In later sections, this report gives further background on research in reactive and incremental computing (Section 4), and further details on the new IC Workshop (Section 5).

Acknowledgments. Co-organizing this seminar with Camil, Sebastian and Shriram was a pleasure. I am especially thankful to Shriram for organizing the event's structure, and moderating group discussions and group decision making during its execution. We organizers are all thankful to the participants, who all brought a unique insight to the seminar, which in my humble opinion, succeeded in its aims.

2 Table of Contents

| | |
|--|----|
| Executive Summary | |
| <i>Matthew A. Hammer</i> | 1 |
| Event Summary | |
| Poster Sessions | 4 |
| Group discussions | 4 |
| Topic outlines and Wikipedia edits | 5 |
| New workshop on incremental computation | 7 |
| Background | |
| Incremental computing | 7 |
| Reactive programming | 8 |
| Event Outcome: Workshop on Incremental Computation (IC) | |
| <i>Sebastian Erdweg and Matthew A. Hammer</i> | 9 |
| Participants | 12 |

3 Event Summary

The following table summarizes how we organized the three and a half day event. (Monday of that week is a German holiday). Rather than organize the event into talk sessions, we chose more interactive sessions: Posters and introductions (on day 1) and group discussions on the remaining days.

| | | | |
|-------|-----------|---|--|
| Day 1 | Tuesday | <i>First session</i> | <i>Second session</i> |
| | AM | Poster session A | Poster session B |
| | PM | Poster session C | 1-min introductions (1 slide each). |
| Day 2 | Wednesday | <i>First session</i> | <i>Second session</i> |
| | AM | Demos | Group discussion: IC vs RP |
| | PM | Group discussion: Domain-specific | Break: Outside walks |
| Day 3 | Thursday | <i>First session</i> | <i>Second session</i> |
| | AM | Group discussion: Run-time design | Group discussion: Meshing with non-IC/RP |
| | PM | Smaller group discussions: Algorithms, Semantics & Types & Verification | |
| Day 4 | Friday | First session | Second session |
| | AM | Topic outlines, Wikipedia editing: <i>Incremental computing</i> and <i>Reactive programming</i> | |

3.1 Poster Sessions

The following seminar participants presented posters about their research (the organizers evenly distributed themselves among these sessions, indicated in bold):

| | |
|-----------|--|
| Session 1 | Demetrescu , Haller, Khoo, Ley-Wild, Minsky, Salvaneschi, Szábo, Tangwonsan |
| Session 2 | Burckhardt, Cicek and Garg, Erdweg , Hammer , Krishnaswami, Labich, McSherry, Newton, Shah |
| Session 3 | Bhatotia, Courtney, Harkes, Krishnamurthi , Mezini, Pouzet, Shapiro |

3.2 Group discussions

Before and during the seminar, we took surveys of the participants to find topics that for interesting discussions. In the end, the following topics were scheduled into the event.

- Incremental Computing (IC) vs. Reactive Programming (RP).
Our first discussion centered on the differences between the domains of incremental and reactive systems, and I have paraphrased some conclusions from that discussion.
 - In common to both IC and RP, we broadly consider ad hoc approaches “harmful”, in the sense that they lack systematic abstractions, and consequently, may suffer from undefined or inconsistent behavior (“glitches”). The alternative to ad hoc approaches are programming language abstractions and carefully-designed libraries that offer reusable abstractions, with well-defined semantics.
 - In common to both IC and RP, there are common abstractions and implementations based on, e.g., dataflow graphs.
 - Reactive programming, unlike IC, encompasses programs whose behavior interacts with other systems in time, and generally does not terminate.

These computations consist of signal processing, aviotics and control systems, OS kernels, and financial analytics. All of these domains require time-dependent behavior that senses time-dependent inputs. In many cases, there may be real-time constraints. To a first approximation, they lack costly, redundant subcomputations that terminate and repeat over time.

- Incremental computation, unlike RP, is concerned with computational cost. It attempts to *improve the algorithmic efficiency* of computations that terminate, but repeat over time in a changing environment.

The general aim of IC is to improve the asymptotic efficiency of repeated computations, and/or, to cache and reuse large portions of past computations. IC encompasses techniques that make the following tasks more efficient:

- * *data synchronization or versioning*, where the redundant subcomputation to avoid is communication of data that has not changed since the last pass;
- * *program analysis, HTML rendering and spreadsheet evaluation*, where the redundant subcomputation to avoid is the analysis, rendering or calculations that have not been affected by changes since the last pass.

Other discussion topics:

- Domain-specific techniques
- Run-time system design
- Meshing with non-IC/RP
- Semantics & Types: Small, break-out group
- Algorithms: Small, break-out group
- Verification: Small, break-out group

3.3 Topic outlines and Wikipedia edits

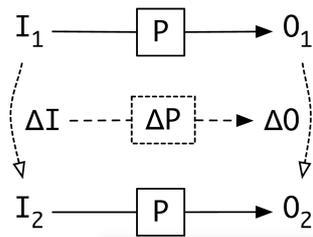
During the final morning of the seminar, the seminar participants brainstormed outlines for new Wikipedia articles on the topics of *Incremental computing* and *Reactive programming*. This exercise forced us to catalog the most important concepts in these fields, and how to best structure their exposition. To record these outlines, we edited Wikipedia collaboratively. The seminar participants broke into two groups, to focus on the incremental and reactive articles in a divide-and-conquer approach.

The outcome of this editing session are new article outlines, with some discussion, and some links to relevant literature. For both articles, much more work is warranted. However, we feel that the consensus reached during the seminar will make future collaborative editing easier at a distance. (e.g., by using the Dagstuhl Seminar email list, or Wikipedia itself, to coordinate).

A recording of the new proposed outlines is included below, for posterity. (Of course, we give permission to Wikipedia to use this outline; in fact, we encourage them to continue to do so!). Below the outlines, we give links to the exact Wikipedia edits.

3.3.1 Article outline: Incremental Computing

- Static Versus Dynamic
- Specialized versus General-Purpose Approaches



■ **Figure 1** Incremental computing of P using ΔP is sound when it is commutative, as above. This diagram shows the relationships between a program P , successive inputs I_1 and I_2 , successive outputs O_1 and O_2 , their relative changes ΔI and ΔO , and the change propagation mechanism that performs ΔO somehow from ΔI .

- Static Methods
 - Program Derivatives
 - View Maintenance
- Dynamic Methods
- Existing systems
 - Compiler and Language Support
 - Frameworks and libraries
- Applications
 - Databases (view maintenance)
 - Build systems
 - Spreadsheets
 - Development Environments
 - Financial Computations
 - Attribute Grammar Evaluation
 - Graph Computations and Queries
 - GUIs (e.g., React and DOM diffing)
 - Scientific applications
- See also
 - Reactive programming
 - Memoization

We also contributed Figure 1, which we used throughout the seminar to orient our group discussions.

3.3.2 Article outline: Reactive programming

- Definition of Reactive Programming
- Approaches to Creating Reactive Programming Languages
 - Dedicated languages that are specific to some domain constraints (such as real-time or embedded computing or hardware description)
 - General-purpose languages that support reactivity
 - Libraries or embedded domain specific languages that enable reactivity alongside or on top of an existing general-purpose programming language
- Programming Models and Semantics
 - Synchrony: is the underlying model of time synchronous versus asynchronous?
 - Determinism: Deterministic versus non-deterministic in both evaluation process and results (the former does not necessarily imply the latter)
 - Update process: callbacks versus dataflow versus actors

- Implementation Challenges
 - Glitches
 - Cyclic Dependencies
 - Interaction with Mutable State
 - Dynamic Updating of the Graph of Dependencies

3.3.3 Wikipedia edits

Our collective edits, relative to the article before our collaborative editing session, are visible at the following Wikipedia URLs:

- **Incremental computing edits:**
https://en.wikipedia.org/w/index.php?title=Incremental_computing&type=revision&diff=743022258&oldid=735698349
- **Reactive programming edits:**
https://en.wikipedia.org/w/index.php?title=Reactive_programming&type=revision&diff=743074812&oldid=740655985

3.4 New workshop on incremental computation

Following the success of this Dagstuhl Seminar, organizers Sebastian Erdweg and Matthew Hammer proposed a new Workshop on Incremental Computing (IC) to PLDI 2017, in Barcelona, Spain. The content of this workshop proposal is included below, in Section 5. PLDI 2017 has since accepted this proposal.

4 Background

4.1 Incremental computing

Incremental computations are those that process input changes faster than a naive re-computation from scratch. Due to the importance and prevalence of incremental, reactive systems, ad hoc variants of incremental computation are ubiquitous in modern software systems. As an everyday example, spreadsheets such as Excel re-calculate selectively based on user interaction and the dynamic dependencies of formula. Incremental computation is needed in this domain, and many others, for efficiency and responsiveness. As other examples, build systems and integrated development environments re-compile selectively, based on the code dependencies. Hence, build systems strive to use a domain-specific form of incremental computation that is aware of compiler dependencies [14, 27]. Doing so is necessary to support interactive development, testing and debugging, which should be responsive to code changes. Meanwhile, the interactive behavior of the visual elements in the development tools, and their interaction with external tools, can be modeled with reactive computation. Finally, modern web browsers (such as Chrome, Firefox, Safari, etc.) house incremental computations that respond to mobile code, whose execution leads to re-computing layout and styling information (viz., dynamic variation of CSS attributes incrementally affect the placement and appearance of modern web pages). These scripts, as well as the browser platform in which they run, are incremental, reactive computations [6].

Due to their prevalence in practical systems used every day, notions of incremental computing abound in computer science broadly, and within research on programming

languages (PL). In the area of PL, researchers are particularly interested in **language-based approaches** to incremental computation. In contrast to the algorithms community that often studies each incremental problem in isolation (e.g., incremental convex hull), PL researchers study large classes of incremental programs that are defined by a general language. Their typical goal is to provide a language and associated technique that is general enough to express the behavior of many incremental or reactive programs. For instance, many general-purpose techniques can derive the incremental behavior of two-dimensional convex hull from the expression of a textbook algorithm for the non-incremental algorithm (e.g., quickhull) [1, 16, 5, 19]. More generally, researchers have shown that for certain algorithms, inputs, and classes of input changes, IC delivers large, even *asymptotic* speed-ups over full reevaluation [4, 2]. IC has been developed in many different language settings [26, 17, 18, 9], and has even been used to address open problems, e.g., in computational geometry [3].

4.2 Reactive programming

Reactive programming languages offer abstractions for processing events generated by dynamic environments. Reactive abstractions encompass both event-driven and data-driven scenarios, relying on graphs to model dependencies in a program. In the former, events are explicitly modeled as a stream generated over time; computations respond to generated events and may trigger further computations along the dependency graph. In the latter, events are modeled as input data changes as in IC frameworks. A data-flow graph describes relationships between objects and changes are automatically propagated throughout the graph.

Similarly to IC environments, a critical aspect in reactive systems is to minimize the amount of recomputations triggered by external discrete events or input data changes. Early examples of event-based reactive environments for real-time systems in embedded software include Signal [15] and Lustre [8].

Functional Reactive Programming (FRP) is a declarative programming model for constructing interactive applications [13, 24, 28]. The chief aim of FRP is to provide a declarative means of specifying programs whose values are time-dependent (stored in signals), whereas the chief aim of IC is to provide time savings for small input changes (stored in special references). The different scope and programming model of FRP makes it hard to imagine using it to write an efficient incremental sorting algorithm, though it may be possible. On the other hand, IC would seem to be an appropriate mechanism for implementing an FRP engine, though the exact nature of this connection remains unclear.

FrTime [10] extends a purely functional subset of PLT Scheme with an instantiation of the FRP paradigm, supporting eager evaluation and *benign impurities* (e.g., imperative commands for drawing and for creating and varying mutable references). The problem of integrating FrTime and object-oriented graphics toolkits has been investigated by [20].

More recently [23] have introduced Flapjax, a reactive extension to JavaScript for Web applications, whose approach is mainly informed by FrTime. Frappé [11] integrates the FRP model with the Java Beans technology, allowing reactive programming in Java. FrTime has also served as a basis for MzTake [22], a scriptable debugger implementing a dataflow language in the tradition of Dalek [25], an earlier programmable debugger that also modeled events using a data-flow graph. SugarCubes [7] and ReactiveML [21] allow reactive programming in Java and OCaml, respectively.

A different data-driven line of research investigates how to mix the reactive paradigm with imperative and object-oriented languages, allowing programmers to declaratively express dataflow constraints between C/C++ objects allocated in a special “reactive memory” heap [12].

5 Event Outcome: Workshop on Incremental Computation (IC)

Sebastian Erdweg (TU Delft, NL) and Matthew A. Hammer (University of Colorado – Boulder, US)

License  Creative Commons BY 3.0 Unported license
© Sebastian Erdweg and Matthew A. Hammer

The content below is from a successful workshop proposal to PLDI 2017. Elsewhere, this proposal referenced the success of this Dagstuhl Seminar as evidence of research community interest.

Due to its cross-cutting nature, research results on and experience with incremental computing is scattered throughout the PL community. The Workshop on Incremental Computing (IC) will provide a platform for researchers and users of incremental computing.

- 4 sessions featuring invited talks from academia and industry as well as contributed talks from the community.
- Type of submission: Contributed talks need to submit talk abstracts, which forms the basis for selection.
- Review process: We will invite a small program committee that selects contributed talks based on the submitted talk abstracts.
- Result dissemination: We will collect talk abstracts from all invited and selected contributed talks and publish them as an openly accessible technical report.

Since incremental computations are cross-cutting PL, we expect significant interest within the PLDI community. Traditionally, static analysis has seen numerous successful applications of incremental computing, and the workshop may spark especial interest in that part of the community. Since this is the first workshop on incremental computing, it is very difficult to estimate the number of attendees; anything between 20 and 80 people seems realistic.

References

- 1 Umut A. Acar, Guy E. Blelloch, Matthias Blume, Robert Harper, and Kanat Tangwongsan. A library for self-adjusting computation. *ENTCS*, 148(2), 2006.
- 2 Umut A. Acar, Guy E. Blelloch, Kanat Tangwongsan, and Duru Türkoğlu. Robust kinetic convex hulls in 3D. In *Proceedings of the 16th Annual European Symposium on Algorithms*, September 2008.
- 3 Umut A. Acar, Andrew Cotter, Benoît Hudson, and Duru Türkoğlu. Dynamic well-spaced point sets. In *Symposium on Computational Geometry*, 2010.
- 4 Umut A. Acar, Alexander Ihler, Ramgopal Mettu, and Özgür Sümer. Adaptive Bayesian inference. In *Neural Information Processing Systems (NIPS)*, 2007.
- 5 Umut A. Acar and Ruy Ley-Wild. Self-adjusting computation with Delta ML. In *Advanced Functional Programming*. Springer Berlin Heidelberg, 2009.
- 6 Brian Anderson, Lars Bergstrom, David Herman, Josh Matthews, Keegan McAllister, Manish Goregaokar, Jack Moffitt, and Simon Sapin. Experience report: Developing

- the servo web browser engine using rust. *CoRR*, abs/1505.07383, 2015. URL: <http://arxiv.org/abs/1505.07383>.
- 7 Frédéric Boussinot and Jean-Ferdys Susini. The SugarCubes Tool Box: a Reactive Java Framework. *Software: Practice and Experience*, 28(14):1531–1550, 1998.
 - 8 P. Caspi, P. Pilaud, N. Halbwachs, and J. Plaice. Lustre, a Declarative Language for Programming Synchronous Systems. In *POPL*, pages 178–188, 1987.
 - 9 Yan Chen, Joshua Dunfield, Matthew A. Hammer, and Umut A. Acar. Implicit self-adjusting computation for purely functional programs. *J. Functional Programming*, 24(1):56–112, 2014.
 - 10 Gregory H. Cooper and Shriram Krishnamurthi. Embedding dynamic dataflow in a call-by-value language. In *ESOP*, 2006.
 - 11 Antony Courtney. Frappé: Functional Reactive Programming in Java. In *PADL*, pages 29–44, 2001.
 - 12 Camil Demetrescu, Irene Finocchi, and Andrea Ribichini. Reactive imperative programming with dataflow constraints. *ACM Trans. Program. Lang. Syst.*, 37(1):3:1–3:53, 2014.
 - 13 Conal Elliott and Paul Hudak. Functional Reactive Animation. In *ICFP*, pages 263–273, 1997.
 - 14 Sebastian Erdweg, Moritz Lichter, and Manuel Weiel. A sound and optimal incremental build system with dynamic dependencies. In *OOPSLA’15*, pages 89–106. ACM, 2015.
 - 15 P. Le Guernic, A. Benveniste, P. Bournai, and T. Gautier. SIGNAL – A Data Flow-Oriented Language for Signal Processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(2):362–374, 1986.
 - 16 Matthew Hammer and Umut A. Acar. Memory management for self-adjusting computation. In *ISMM*, 2008.
 - 17 Matthew Hammer, Umut A. Acar, Mohan Rajagopalan, and Anwar Ghuloum. A proposal for parallel self-adjusting computation. In *DAMP’07: Declarative Aspects of Multicore Programming*, 2007.
 - 18 Matthew A. Hammer, Umut A. Acar, and Yan Chen. CEAL: a C-based language for self-adjusting computation. In *ACM SIGPLAN Conference on Programming Language Design and Implementation*, 2009.
 - 19 Matthew A. Hammer, Joshua Dunfield, Kyle Headley, Nicholas Labich, Jeffrey S. Foster, Michael Hicks, and David Van Horn. Incremental computation with names (extended version). arXiv:1503.07792 [cs.PL], 2015.
 - 20 Daniel Ignatoff, Gregory H. Cooper, and Shriram Krishnamurthi. Crossing State Lines: Adapting Object-Oriented Frameworks to Functional Reactive Languages. In *FLOPS*, pages 259–276, 2006.
 - 21 Louis Mandel and Marc Pouzet. ReactiveML, a Reactive Extension to ML. In *PPDP*, pages 82–93, 2005.
 - 22 Guillaume Marceau, Gregory H. Cooper, Jonathan P. Spiro, Shriram Krishnamurthi, and Steven P. Reiss. The design and implementation of a dataflow language for scriptable debugging. *Automated Software Engg.*, 14(1):59–86, 2007.
 - 23 Leo A. Meyerovich, Arjun Guha, Jacob Baskin, Gregory H. Cooper, Michael Greenberg, Aleks Bromfield, and Shriram Krishnamurthi. Flapjax: a Programming Language for Ajax Applications. In *OOPSLA*, pages 1–20, 2009.
 - 24 Henrik Nilsson, Antony Courtney, and John Peterson. Functional reactive programming, continued. In *Proceedings of the 2002 ACM SIGPLAN Haskell Workshop (Haskell’02)*, pages 51–64, Pittsburgh, Pennsylvania, USA, October 2002. ACM Press.
 - 25 Ronald A. Olsson, Richard H. Crawford, and W. Wilson Ho. A dataflow approach to event-based debugging. *Softw. Pract. Exper.*, 21(2):209–229, 1991.

- 26 Ajeet Shankar and Rastislav Bodik. DITTO: Automatic incrementalization of data structure invariant checks (in Java). In *Programming Language Design and Implementation*, 2007.
- 27 Tamás Szabó, Sebastian Erdweg, and Markus Völter. IncA: A DSL for the definition of incremental program analyses. In *Proceedings of International Conference on Automated Software Engineering (ASE)*. ACM, 2016.
- 28 Zhanyong Wan and Paul Hudak. Functional Reactive Programming from First Principles. In *PLDI*, pages 242–252, 2000.

Participants

- Pramod Bhatotia
TU Dresden, DE
- Sebastian Burckhardt
Microsoft Research –
Redmond, US
- Ezgi Cicek
MPI-SWS – Saarbrücken, DE
- Antony Courtney
San Francisco, US
- Camil Demetrescu
Sapienza University of Rome, IT
- Sebastian Erdweg
TU Delft, NL
- Deepak Garg
MPI-SWS – Saarbrücken, DE
- Philipp Haller
KTH Royal Institute of
Technology – Stockholm, SE
- Matthew A. Hammer
University of Colorado –
Boulder, US
- Daco Harkes
TU Delft, NL
- Kyle Headley
University of Colorado –
Boulder, US
- Yit Phang Khoo
The MathWorks Inc. –
Natick, US
- Shriram Krishnamurthi
Brown University –
Providence, US
- Neel Krishnaswami
University of Cambridge, GB
- Nicholas Labich
University of Maryland –
College Park, US
- Ruy Ley-Wild
LogicBlox – Atlanta, US
- Frank McSherry
Richmond, US
- Mira Mezini
TU Darmstadt, DE
- Yaron Minsky
Jane Street – New York, US
- Ryan R. Newton
Indiana University –
Bloomington, US
- Marc Pouzet
ENS – Paris, FR
- Guido Salvaneschi
TU Darmstadt, DE
- Rohin Shah
University of California –
Berkeley, US
- R. Benjamin Shapiro
University of Colorado –
Boulder, US
- Tamás Szabó
TU Delft, NL
- Kanat Tangwongsan
Mahidol University, TH



Algebraic and Combinatorial Methods in Computational Complexity

Edited by

Valentine Kabanets¹, Thomas Thierauf², Jacobo Tóran³, and Christopher Umans⁴

1 Simon Fraser University, CA, kabanets@cs.sfu.ca

2 Aalen University, DE, thomas.thierauf@uni-ulm.de

3 Ulm University, DE, jacobito@uni-ulm.de

4 CalTech – Pasadena, US, umans@cs.caltech.edu

Abstract

Computational Complexity is concerned with the resources that are required for algorithms to detect properties of combinatorial objects and structures. It has often proven true that the best way to argue about these combinatorial objects is by establishing a connection (perhaps approximate) to a more well-behaved algebraic setting. Indeed, many of the deepest and most powerful results in Computational Complexity rely on algebraic proof techniques. The Razborov-Smolensky polynomial-approximation method for proving constant-depth circuit lower bounds, the PCP characterization of NP, and the Agrawal-Kayal-Saxena polynomial-time primality test are some of the most prominent examples.

The algebraic theme continues in some of the most exciting recent progress in computational complexity. There have been significant recent advances in algebraic circuit lower bounds, and the so-called chasm at depth 4 suggests that the restricted models now being considered are not so far from ones that would lead to a general result. There have been similar successes concerning the related problems of polynomial identity testing and circuit reconstruction in the algebraic model (and these are tied to central questions regarding the power of randomness in computation).

Another surprising connection is that the algebraic techniques invented to show lower bounds now prove useful to develop efficient algorithms. For example, Williams showed how to use the polynomial method to obtain faster all-pair-shortest-path algorithms. This emphasizes once again the central role of algebra in computer science.

The seminar aims to capitalize on recent progress and bring together researchers who are using a diverse array of algebraic methods in a variety of settings. Researchers in these areas are relying on ever more sophisticated and specialized mathematics and this seminar can play an important role in educating a diverse community about the latest new techniques, spurring further progress.

Seminar Oktober 9–14, 2016 – <http://www.dagstuhl.de/16411>

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes, F.2 Analysis of Algorithms and Problem Complexity.

Keywords and phrases Computational Complexity, lower bounds, approximation, pseudo-randomness, derandomization, circuits

Digital Object Identifier 10.4230/DagRep.6.10.13



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

Algebra in Computational Complexity, *Dagstuhl Reports*, Vol. 6, Issue 10, pp. 13–32

Editors: Valentine Kabanets, Thomas Thierauf, Jacobo Tóran, and Christopher Umans



DAGSTUHL
REPORTS

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

1 Executive Summary

Valentine Kabanets

Thomas Thierauf

Jacobo Tóran

Christopher Umans

License  Creative Commons BY 3.0 Unported license

© Valentine Kabanets, Thomas Thierauf, Jacobo Tóran, and Christopher Umans

The seminar brought together more than 40 researchers covering a wide spectrum of complexity theory. The focus on algebraic methods showed the great importance of such techniques for theoretical computer science. We had 25 talks, most of them lasting about 40 minutes, leaving ample room for discussions. In the following we describe the major topics of discussion in more detail.

Circuit Complexity

This is an area of fundamental importance to Complexity. Circuit Complexity was one of the main topics in the seminar. Still it remains a big challenge to prove strong upper and lower bounds. Also Polynomial Identity Testing (PIT) plays a central role.

The seminar started with a talk by *Steve Fenner*. In a breakthrough result, he showed how to solve the perfect matching problem in bipartite graphs (almost) efficiently in parallel, by circuits of quasi-polynomial size and $O(\log^2 n)$ depth (in quasi-NC). This solves a problem open since more than 30 years. *Rohit Gurjar* showed how to extend the result even further to *linear matroid intersection*, where bipartite perfect matching is a special case of.

Both of the above results can be read as a singularity test of certain symbolic matrices. We had several talks dealing with determining singularity or computing the rank of a symbolic matrix. *Rafael Oliveira* presented an efficient algorithm for the symbolic singularity problem in the *non-commutative* setting. In the *commutative* setting, the complexity is a major open problem. Many other important problems reduce to it. *Markus Bläser* presented an *approximation algorithm* (PTAS) for the rank of a symbolic matrix. Surprisingly, this is achieved with a greedy-algorithm. *Kristoffer Hansen* showed a different kind of approximation for low rank binary matrices.

We have seen some great work on *Polynomial Identity Testing* (PIT) and circuit lower bounds recently, in particular on depth-3 and depth 4 circuits, and on arithmetic branching programs, which has brought us very close to statements that are known to imply $VP \neq VNP$, the analogue of the P vs. NP question in the arithmetic world. With respect to PIT, an ambitious goal is to come up with a hitting set construction for a specific model. A hitting set is a set of instances such that every non-zero polynomial in the model has a non-root in the set. This would solve the PIT problem in the *black box* model.

PIT is known to be efficiently solvable by *randomized* algorithms, for example when we consider arithmetic circuits. Things get a bit different when we consider *noncommutative* circuits. Now the standard test cannot be directly applied because the polynomials can have exponential degree, and hence doubly exponentially many monomials. *V. Arvind* presented a randomized polynomial identity test for noncommutative arithmetic circuits for the case when the polynomial has only exponentially many monomials.

One of the most successful methods for proving lower bounds for arithmetic circuits is to consider the dimension of the span of the *partial derivatives* of a polynomial. *Pascal Koiran*

considered the complexity of the problem to compute this dimension. He showed that it is $\#P$ -hard. It remained open whether the problem is $\#P$ -complete.

Another important notion when proving lower bounds is the *algebraic independence* of arithmetic circuits. In 2015, Kumar and Saraf presented lower bounds and hitting sets for a class of depth-4 circuits that have low algebraic rank. Unfortunately, their technique requires base fields of characteristic zero, or at least exponentially large characteristic. *Nitin Saxena* closed this gap and showed how to make the approach work over *every* field.

Michael Forbes showed that lower bounds for certain algebraic circuits imply lower bounds in proof complexity.

Or Meir talked on one of the major open problems in complexity theory: proving super-polynomial lower bounds on the size of formulas. Karchmer, Raz, and Wigderson suggested an approach to this problem. The *KRW-conjecture* states that the formula complexity of two functions f and g roughly adds up when we consider the composed function $g \circ f$. They showed that the conjecture implies super-polynomial formula lower bounds. In his talk, Or Meir did a step to prove the conjecture: he proved a special case, namely when f is the parity-function. His proof uses techniques from communication complexity.

Valiant introduced the arithmetic analogue of classes P and NP. Very roughly, the class VP contains all multivariate polynomials that can be computed (non-uniformly) by polynomial-size arithmetic circuits, and the class VNP contains all multivariate polynomials that have coefficients computable by VP-circuits. The question whether VP is different from VNP plays the role of the P-NP question in algebraic complexity theory. Valiant showed that the permanent is complete for VNP. But for VP, only artificially constructed functions were known to be complete. In her talk, *Meena Mahajan* described several polynomial families complete for VP and for VNP, based on the notion of graph homomorphism polynomials.

Complexity

Since the famous AKS-primality test, prime numbers can be recognized efficiently. The *construction* of prime numbers is still a challenging task. The best known deterministic algorithms have only exponential running time. *Rahul Santhanam* presented a randomized subexponential time algorithm that outputs primes, and only primes, with high probability, and moreover, the output is mostly the same prime. This is called a *zero-error pseudo-deterministic* algorithm.

Since the famous Isolation Lemma of Mulmuley, Vazirani, Vazirani, researchers recognized the power of isolation. For example, the bipartite perfect matching and the matroid intersection algorithms mentioned above, both rely on isolating a minimum weight solution, *Nutan Limaye* studied the problem of isolating an s - t -path in a directed graph. She proved that a randomized logspace algorithm that isolates such a path can be used to show $NL \subseteq L/poly$.

Derandomization is an area where there are tight connections between lower bounds and algorithms. Strong enough circuit lower bounds can be used to construct pseudo-random generators that can then be used to simulate randomized algorithms with only polynomial overhead. The polynomial overhead is fine for algorithms running in polynomial time. However, in case of subexponential randomized algorithms, this overhead makes the resulting deterministic algorithm more or less useless. *Ronen Shaltiel* showed how to overcome this problem by achieving a more modest overhead. He needs, however, stronger lower bounds to begin with. Further talks on pseudo-random generators and randomness extractors were given by *Amnon Ta-Shma* and *William Hoza*.

Chris Umans gave an evening talk presenting a recent breakthrough in additive combinatorics, the resolution of the so-called *cap-set conjecture* by Ellenberg and Gijswijt. This result has implications for the Cohn-Umans group-theoretic approach for matrix multiplication, and elsewhere in Complexity.

Coding Theory

Error-correcting codes, particularly those constructed from polynomials, i.e. Reed-Solomon codes or Reed-Muller codes, lie at the heart of many significant results in Computational Complexity. *Shubhangi Saraf* gave a talk on locally-correctable and locally-testable codes. *Swastik Kopparty* generalized the well known decoding algorithm for Reed-Solomon codes to higher dimensions. He presented an efficient algorithm to decode Reed-Muller codes when the evaluation points are an arbitrary product set S^m , for some m , when S is larger than the degree of the polynomials.

Quantum Complexity

Complexity issues arising in the context of quantum computation are an important area in complexity theory since several decades. In the workshop, we had two talks related to quantum complexity. *Farid Ablayev* talked about the notion of quantum hash function and how to construct such functions. He also explained some of its applications for constructing quantum message authentication codes. *Ryan O'Donnell* explained about the *quantum tomography problem* and how this special case of *quantum spectrum estimation* can be solved combinatorially by understanding certain statistics of random words.

Conclusion

As is evident from the list above, the talks ranged over a broad assortment of subjects with the underlying theme of using algebraic and combinatorial techniques. It was a very fruitful meeting and has hopefully initiated new directions in research. Several participants specifically mentioned that they appreciated the particular focus on a common class of *techniques* (rather than end results) as a unifying theme of the workshop. We look forward to our next meeting!

2 Table of Contents

Executive Summary

Valentine Kabanets, Thomas Thierauf, Jacobo Tóran, and Christopher Umans . . . 14

Overview of Talks

| | |
|---|----|
| Quantum Fingerprinting and Quantum Hashing. Computational and Cryptographical Aspects <i>Farid Ablayev</i> | 19 |
| PIT for noncommutative circuits <i>Vikraman Arvind</i> | 20 |
| A deterministic PTAS for the commutative rank of a symbolic matrix <i>Markus Bläser</i> | 20 |
| Non-algebraic methods for (and against) secret sharing <i>Andrej Bogdanov</i> | 21 |
| An Efficient Deterministic Simulation Theorem <i>Arkadev Chattopadhyay</i> | 21 |
| Bipartite Perfect Matching is in quasi-NC <i>Stephen A. Fenner</i> | 21 |
| Proof Complexity Lower Bounds from Algebraic Circuit Complexity <i>Michael Forbes</i> | 22 |
| Linear Matroid Intersection is in quasi-NC <i>Rohit Gurjar</i> | 23 |
| On Low Rank Approximation of Binary Matrices <i>Kristoffer Arnsfelt Hansen</i> | 23 |
| Targeted Pseudorandom Generators, Simulation Advice Generators, and Derandomizing Logspace <i>William Hoza</i> | 24 |
| The complexity of partial derivatives <i>Pascal Koiran</i> | 24 |
| Decoding Reed-Muller codes over product sets <i>Swastik Kopparty</i> | 25 |
| Lower Bounds for Elimination via Weak Regularity <i>Michal Koucký</i> | 25 |
| Isolation Lemma for Directed Reachability and NL vs. L <i>Nutan Limaye</i> | 26 |
| Enumerator polynomials: Completeness and Intermediate Complexity <i>Meena Mahajan</i> | 26 |
| Toward the KRW Composition Conjecture: Cubic Formula Lower Bounds via Communication Complexity <i>Or Meir</i> | 27 |
| Efficient Quantum Tomography and Longest Increasing Subsequences <i>Ryan O'Donnell</i> | 27 |

| | |
|---|----|
| Operator Scaling and Applications to Algebraic Complexity, Mathematics and Optimization <i>Rafael Oliveira</i> | 28 |
| On the Complexity of Generating Primes <i>Rahul Santhanam</i> | 28 |
| High rate locally-correctable and locally-testable codes with sub-polynomial query complexity <i>Shubhangi Saraf</i> | 29 |
| Algebraic independence over positive characteristic: New criterion and applications to locally low algebraic rank circuits <i>Nitin Saxena</i> | 29 |
| Pseudorandomness when the odds are against you <i>Ronen Shaltiel</i> | 30 |
| Explicit two-source extractors for near-logarithmic min-entropy <i>Amnon Ta-Shma</i> | 30 |
| A one-sided error randomized protocol for Gap Hamming Distance problem <i>Nikolay K. Vereshchagin</i> | 31 |
| Pointer chasing via triangular discrimination <i>Amir Yehudayoff</i> | 31 |
| Participants | 32 |

3 Overview of Talks

3.1 Quantum Fingerprinting and Quantum Hashing. Computational and Cryptographical Aspects

Farid Ablyayev (Kazan State University, RU)

License © Creative Commons BY 3.0 Unported license
© Farid Ablyayev

Joint work of Farid Ablyayev, Marat Ablyayev, Alexander Vasiliev, Mansur Ziatdinov

Main reference F. Ablyayev, M. Ablyayev, A. Vasiliev, and M. Ziatdinov, “Quantum Fingerprinting and Quantum Hashing. Computational and Cryptographical Aspects”, *Baltic J. Modern Computing*, Vol. 4(4), pp. 860–875, 2016.

URL <http://dx.doi.org/10.22364/bjmc.2016.4.4.17>

Rusins Freivalds was one of the first researchers who introduced methods (later called fingerprinting) for constructing effective randomized algorithms (which are more effective than any deterministic algorithm) (Freivalds, 1977, 1979). In quantum case, fingerprinting is a procedure that maps classical data to a quantum state that identifies the original data (with high probability). One of the first applications of the quantum fingerprinting method is due to Ambainis and Freivalds (1998): for a specific language they have constructed a quantum finite automaton with an exponentially smaller size than any classical randomized automaton. An explicit definition of the quantum fingerprinting was introduced by Buhrman et al. in (2001) for constructing effective quantum communication protocol for equality testing.

We define a notion of quantum hash function which is quantum one-way function and quantumly collision resistant function. We show that one-way property and collision resistance property are correlated for a quantum hash function. The more the function is one-way the less it is collision resistant and vice versa. We show that such a correlation can be balanced.

We present an approach for quantum hash function constructions by establishing a connection with small biased sets (Naor & Naor, 1990) and quantum hash function constructions: we prove that small sized ϵ -biased sets allow to generate balanced quantum hash functions. Such a connection adds to the long list of small-biased sets? applications. In particular it was observed in (Naor & Naor, 1990; Ben-Sasson et al., 2003) that the ϵ -bias property is closely related to the error-correcting properties of linear codes. Note that the quantum fingerprinting function from (Buhrman et al., 2001) is based on a binary error-correcting code and so it solves the problem of constructing quantum hash functions for the binary case. For the general case, ϵ -bias does not correspond to Hamming distance. Thus, in contrary to the binary case, an arbitrary linear error correcting code cannot be used directly for quantum hash functions.

Next, recall that any ϵ -biased set gives rise to a Cayley expander graph (Alon & Roichman, 1994). We show how such graphs generate balanced quantum hash functions. Every expander graph can be converted to a bipartite expander graph. The generalization of these bipartite expander graphs is the notion of extractor graphs. Such point of view gives a method for constructing quantum hash functions based on extractors.

This construction of quantum hash functions is applied to define the notion of keyed quantum hash functions. The latter is used for constructing quantum hash-based message authentication codes (QMAC). The security proof of QMAC is based on using strong extractors against quantum storage developed by Ta-Shma (2009).

3.2 PIT for noncommutative circuits

Vikraman Arvind (*The Institute of Mathematical Sciences, IN*)

License  Creative Commons BY 3.0 Unported license
© Vikraman Arvind

Joint work of Vikraman Arvind, Partha Mukhopadhyay, S. Raja

Main reference V. Arvind, P. Mukhopadhyay, S. Raja, “Randomized Polynomial Time Identity Testing for Noncommutative Circuits”, ECCC TR16-89, 2016.

URL <http://eccc.hpi-web.de/report/2016/089/>

In this talk we show that the black-box polynomial identity testing for noncommutative polynomials $f \in \mathbb{F}\langle z_1, z_2, \dots, z_n \rangle$ of degree D and sparsity t , can be done in randomized $\text{poly}(n, \log t, \log D)$ time. As a consequence, if the black-box contains a circuit C of size s computing $f \in \mathbb{F}\langle z_1, z_2, \dots, z_n \rangle$ which has at most t non-zero monomials, then the identity testing can be done by a randomized algorithm with running time polynomial in s and n and $\log t$. This makes significant progress on a question that has been open for over ten years.

The earlier result by Bogdanov and Wee [BW05], using the classical Amitsur-Levitski theorem, gives a randomized polynomial-time algorithm only for circuits of polynomially bounded syntactic degree. In our result, we place no restriction on the degree of the circuit.

Our algorithm is based on automata-theoretic ideas introduced in [AMS08,AM08]. In those papers, the main idea was to construct deterministic finite automata that isolate a single monomial from the set of nonzero monomials of a polynomial f in $\mathbb{F}\langle z_1, z_2, \dots, z_n \rangle$. In the present paper, since we need to deal with exponential degree monomials, we carry out a different kind of monomial isolation using nondeterministic automata.

3.3 A deterministic PTAS for the commutative rank of a symbolic matrix

Markus Bläser (*Universität des Saarlandes, DE*)

License  Creative Commons BY 3.0 Unported license
© Markus Bläser

Joint work of Markus Bläser, Gorav Jindal, Anurag Pandey

Main reference M. Bläser, G. Jindal, A. Pandey, “Greedy Strikes Again: A Deterministic PTAS for Commutative Rank of Matrix Spaces”, ECCC TR16-145, 2016.

URL <http://eccc.hpi-web.de/report/2016/145/>

We present a deterministic PTAS for computing the commutative rank of a symbolic matrix or equivalently, of a given matrix space B . More specifically, given a matrix space $B \subseteq F^{n \times n}$ and a rational number $\epsilon > 0$, we give an algorithm that runs in time $O(n^{4+3/\epsilon})$ and computes a matrix A such that the rank of A is at least $(1 - \epsilon)$ times the commutative rank of B . The algorithm is the natural greedy algorithm. It always takes the first set of k matrices that will increase the rank of the matrix constructed so far until it does not find any improvement, where the size of the set k depends on ϵ .

3.4 Non-algebraic methods for (and against) secret sharing

Andrej Bogdanov (The Chinese University of Hong Kong, HK)

License © Creative Commons BY 3.0 Unported license
© Andrej Bogdanov

When we talk about secret sharing things that come to mind are algebraic objects like finite fields, polynomials, codes, etc. We take on a probabilistic viewpoint and use analytic, combinatorial, and game-theoretic tools to rediscover some old connections and answer questions about the complexity of secret sharing and prove new lower bounds on share size in threshold schemes.

3.5 An Efficient Deterministic Simulation Theorem

Arkadev Chattopadhyay (Tata Institute of Fundamental Research – Mumbai, IN)

License © Creative Commons BY 3.0 Unported license
© Arkadev Chattopadhyay
Joint work of Arkadev Chattopadhyay, Michal Koucky, Bruno Loff, Sagnik Mukhopadhyay

Recently, proving theorems of the form that the communication complexity of a composed function $f \circ g$ is essentially of the order of the decision tree complexity of f times the communication complexity of g has received a lot of attention. In particular, Goos-Pitassi-Watson (2015) simplified the proof of such a theorem for deterministic complexity due to Raz-McKenzie (1997) that worked only when g is the Indexing function. They used this theorem to settle a longstanding open problem in communication complexity. The Raz-McKenzie theorem needs the size of the Indexing gadget to be at least n^{20} , where n is the number of instances of Index.

We identify a simple sufficient condition for g to be satisfied to prove such deterministic simulation theorems. Using this, we show that $CC(f \circ IP) = \Omega(DT(f) \cdot m)$, provided $m = \Omega(\log n)$, where IP is the inner-product function. This gives an exponential improvement over the gadget size of Raz and McKenzie.

3.6 Bipartite Perfect Matching is in quasi-NC

Stephen A. Fenner (University of South Carolina, US)

License © Creative Commons BY 3.0 Unported license
© Stephen A. Fenner
Joint work of Stephen A. Fenner, Rohit Gurjar, Thomas Thierauf
Main reference S. A. Fenner, R. Gurjar, T. Thierauf, “Bipartite Perfect Matching is in quasi-NC”, ECCC TR16-177, 2015.
URL <http://eccc.hpi-web.de/report/2015/177/>

We show that the bipartite perfect matching problem is in QuasiNC². That is, it has uniform circuits of quasi-polynomial size $n^{O(\log n)}$, and $O(\log^2 n)$ depth. Previously, only an exponential upper bound was known on the size of such circuits with poly-logarithmic depth.

We obtain our result by an almost complete derandomization of the famous Isolation Lemma when applied to yield an efficient randomized parallel algorithm for the bipartite perfect matching problem.

3.7 Proof Complexity Lower Bounds from Algebraic Circuit Complexity

Michael Forbes (Stanford University, US)

License  Creative Commons BY 3.0 Unported license
© Michael Forbes

Joint work of Michael Forbes, Amir Shpilka, Iddo Zameret, Avi Wigderson

Main reference M. Forbes, A. Shpilka, I. Zameret, A. Wigderson, “Proof Complexity Lower Bounds from Algebraic Circuit Complexity”, ECCC TR16-98, 2016.

URL <http://eccc.hpi-web.de/report/2016/098/>

Proof complexity studies the complexity of mathematical proofs, with the aim of exhibiting (true) statements whose proofs are always necessarily long. One well-known proof system is Hilbert’s Nullstellensatz, which shows that if the family $F = \{f_1, \dots, f_m\}$ of n -variate polynomials have no common solution to the system $f_1 = \dots = f_m = 0$, then there is a proof of this fact of the following form: there are polynomials $G = \{g_1, \dots, g_m\}$ such that $f_1 \cdot g_1 + \dots + f_m \cdot g_m = 1$ is an identity. From the perspective of computer science, it is most natural to assume that the *boolean axioms* $x_i^2 - x_i$ are among the polynomials F , and to ask how succinctly one can express the proof G . Assuming $NP \neq coNP$, there must be systems F such that any proof G requires super-polynomial size to write down, and the goal is to furnish such systems F unconditionally.

Substantial work on the Nullstellensatz system has measured the complexity of G in terms of their degree or sparsity, and obtained the desired lower bounds for these measures. Grochow and Pitassi have recently argued that it is natural to measure the complexity of G by the size needed to express them as algebraic circuits, as this can be exponentially more succinct than counting monomials. They called the resulting system the Ideal Proof System (IPS), and showed that it captures the power of well-known strong proof systems such as the Frege proof system, as well as showing that certain natural lower bounds for the size of IPS proofs would imply $VP \neq VNP$, an algebraic analogue of $P \neq NP$. This is in contrast to other proof systems, where direct ties to computational lower bounds are often lacking.

Motivated by their work, we study the IPS proof system further. We first show that weak subsystems of IPS can be quite powerful. We consider the *subset-sum axiom*, that $x_1 + \dots + x_n + 1$ is unsatisfiable over the boolean cube. In prior work, Impagliazzo, Pudlak, and Sgall showed that any proof of unsatisfiability requires exponentially many monomials to write down. Here, we give an efficient proof even in restricted subclasses of the IPS proof system, showing that the proof can be expressed as a polynomial-size read-once oblivious algebraic branching program (roABP) or depth-3 multilinear formula.

We then consider lower bounds for subclasses of IPS. We obtain certain extensions to existing lower bound techniques, obtaining *functional lower bounds* as well as *lower bounds for multiples*. Using these extensions, we show that variants of the subset-sum axiom require super-polynomially large proofs to prove their unsatisfiability when the size of the algebraic circuits are measured as roABPs, sums of powers of low-degree polynomials, or multilinear formulas.

3.8 Linear Matroid Intersection is in quasi-NC

Rohit Gurjar (Aalen University, DE)

License © Creative Commons BY 3.0 Unported license
© Rohit Gurjar

Joint work of Rohit Gurjar, Thomas Thierauf

Given two matroids on the same ground set, their intersection is the collection of common independent sets. Matroid intersection problem asks to find the maximum cardinality of a common independent set. The problem is in P [Edmonds], and is in RNC for linear matroids [Lovász]. The RNC algorithm is via the isolation lemma, which we derandomize to get a quasi-NC algorithm.

Another way to present the result: we get a quasi-polynomial time blackbox identity testing for the family of polynomials computed by $\det(A_1 z_1 + A_2 z_2 + \dots + A_m z_m)$, where A_i 's are rank 1 matrices.

3.9 On Low Rank Approximation of Binary Matrices

Kristoffer Arnsfelt Hansen (Aarhus University, DK)

License © Creative Commons BY 3.0 Unported license
© Kristoffer Arnsfelt Hansen

Joint work of Chen Dan, Kristoffer Arnsfelt Hansen, He Jiang, Liwei Wang, Yuchen Zhou

Main reference C. Dan, K. A. Hansen, H. Jiang, L. Wang, Y. Zhou, "On Low Rank Approximation of Binary Matrices", arXiv:1511.01699v1 [cs.CC], 2015.

URL <https://arxiv.org/abs/1511.01699v1>

We consider the problem of low rank approximation of binary matrices. Here we are given a $d \times n$ binary matrix \mathbf{A} and a small integer $k < d$. The goal is to find two binary matrices \mathbf{U} and \mathbf{V} of sizes $d \times k$ and $k \times n$ respectively, so that the Frobenius norm of $\mathbf{A} - \mathbf{UV}$ is minimized. There are two models of this problem, depending on the definition of the product of binary matrices: The GF(2) model and the Boolean semiring model. Previously, the only known results are 2-approximation algorithms for the special case $k = 1$ (where the two models are equivalent).

In this paper, we present algorithms for GF(2) and Boolean models respectively. For the GF(2) model, we give a $(\frac{k}{2} + 1 + \frac{k}{2(2^k - 1)})$ -approximation algorithm, which runs in time $O(dn^{k+1})$. For $k = 1$, the approximation ratio is 2. For the Boolean model, we give an algorithm which achieves $(2^{k-1} + 1)$ -approximation and runs in time $O((2^k + 2)!n^{2^k}d)$. We also show that the low rank binary matrix approximation problem is NP-hard for $k = 1$.

3.10 Targeted Pseudorandom Generators, Simulation Advice Generators, and Derandomizing Logspace

William Hoza (University of Texas – Austin, US)

License  Creative Commons BY 3.0 Unported license
© William Hoza

Joint work of William Hoza, Chris Umans

Main reference W. M. Hoza, C. Umans, “Targeted Pseudorandom Generators, Simulation Advice Generators, and Derandomizing Logspace”, arXiv:1610.01199v3 [cs.CC], 2016.

URL <https://arxiv.org/abs/1610.01199v3>

We consider two generalizations of the concept of a pseudorandom generator against logspace. A targeted pseudorandom generator against logspace takes as input a short uniform random seed and a finite automaton; it outputs a long bitstring which looks random to that particular automaton. (Targeted pseudorandom generators were introduced by Goldreich in the BPP setting.) A simulation advice generator for logspace stretches a small uniform random seed into a long advice string; the requirement is that there is some logspace algorithm which, given a finite automaton and this advice string, simulates the automaton reading a long uniform random input. We prove that the intersection over all $\alpha > 0$ of $\text{promise-BPSPACE}(\log^{1+\alpha} n)$ is equal to the corresponding deterministic class if and only if every targeted pseudorandom generator against logspace can be transformed into a comparable simulation advice generator for logspace. In particular, if every derandomization of logspace yields a comparable (ordinary) pseudorandom generator, then BPL is contained in $\text{DSPACE}(\log^{1+\alpha} n)$ for every $\alpha > 0$. We also observe that in the uniform setting, targeted pseudorandom generators against logspace can be transformed into comparable simulation advice generators.

3.11 The complexity of partial derivatives

Pascal Koiran (ENS – Lyon, FR)

License  Creative Commons BY 3.0 Unported license
© Pascal Koiran

Joint work of Pascal Koiran, Ignacio Garcia-Marco, Timothée Pecatte, Stéphan Thomassé

Main reference P. Koiran, I. Garcia-Marco, T. Pecatte, S. Thomassé, “On the complexity of partial derivatives”, arXiv:1607.05494v1 [cs.CC], 2016.

URL <https://arxiv.org/abs/1607.05494v1>

The method of partial derivatives is one of the most successful lower bound methods for arithmetic circuits. It uses as a complexity measure the dimension of the span of the partial derivatives of a polynomial. In this paper, we consider this complexity measure as a computational problem: for an input polynomial given as the sum of its nonzero monomials, what is the complexity of computing the dimension of its space of partial derivatives?

We show that this problem is $\#P$ -hard and we ask whether it belongs to $\#P$. We analyze the *trace method*, recently used in combinatorics and in algebraic complexity to lower bound the rank of certain matrices. We show that this method provides a polynomial-time computable lower bound on the dimension of the span of partial derivatives, and from this method we derive closed-form lower bounds. We leave as an open problem the existence of an approximation algorithm with reasonable performance guarantees.

3.12 Decoding Reed-Muller codes over product sets

Swastik Kopparty (Rutgers University, US)

License © Creative Commons BY 3.0 Unported license
© Swastik Kopparty

Joint work of Swastik Kopparty, John Kim

Main reference J. Y. Kim, S. Kopparty, “Decoding Reed-Muller Codes Over Product Sets”, in Proc. of the 31st Conf. on Computational Complexity (CCC 2016), LIPICs, Vol. 50, pp. 11:1–11:28, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.

URL <http://dx.doi.org/10.4230/LIPIcs.CCC.2016.11>

We give a polynomial time algorithm to decode multivariate polynomial codes of degree d up to half their minimum distance, when the evaluation points are an arbitrary product set S^m , for every $d < |S|$. Previously known algorithms can achieve this only if the set S has some very special algebraic structure, or if the degree d is significantly smaller than $|S|$. We also give a near-linear time randomized algorithm, which is based on tools from list-decoding, to decode these codes from nearly half their minimum distance, provided $d < (1 - \epsilon)|S|$ for constant $\epsilon > 0$.

Our result gives an m -dimensional generalization of the well known decoding algorithms for Reed-Solomon codes, and can be viewed as giving an algorithmic version of the Schwartz-Zippel lemma.

3.13 Lower Bounds for Elimination via Weak Regularity

Michal Koucký (Charles University, CZ)

License © Creative Commons BY 3.0 Unported license
© Michal Koucký

Joint work of Michal Koucký, Arkadev Chattopadhyay, Pavel Dvorak, Bruno Loff, and Sagnik Mukhopadhyay

Main reference A. Chattopadhyay, P. Dvorak, M. Koucký, B. Loff, S. Mukhopadhyay, “Lower Bounds for Elimination via Weak Regularity”, ECCCC’ TR16-156, 2016.

URL <http://ecccc.hpi-web.de/report/2016/165/>

We consider the problem of elimination in communication complexity, that was first raised by Ambainis et al. (2001) and later studied by Beimel et al. (2014) for its connection to the famous direct sum question. In this problem, let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be any boolean function. Alice and Bob get k inputs x_1, \dots, x_k and y_1, \dots, y_k respectively, with $x_i, y_i \in \{0, 1\}^n$. They want to output a k -bit vector v , such that there exists one index i for which $v_i \neq f(x_i, y_i)$. We prove a general result lower bounding the randomized communication complexity of the elimination problem for f using its discrepancy. Consequently, we obtain strong lower bounds for functions Inner-Product and Greater-Than, that work for exponentially larger values of k than the best previous bounds.

To prove our result, we use a pseudo-random notion called regularity that was first used by Raz and Wigderson (1989). We show that functions with small discrepancy are regular. We also observe that a weaker notion, that we call weak-regularity, already implies hardness of elimination. Finally, we give a different proof, borrowing ideas from Viola (2015), to show that Greater-Than is weakly regular.

3.14 Isolation Lemma for Directed Reachability and NL vs. L

Nutan Limaye (Indian Institute of Technology – Mumbai, IN)

License  Creative Commons BY 3.0 Unported license
© Nutan Limaye

Joint work of Nutan Limaye, Vaibhav Krishan

Main reference V. Krishan, N. Limaye, “Isolation Lemma for Directed Reachability and NL vs. L”, ECCC TR16-155, 2016.

URL <http://eccc.hpi-web.de/report/2016/155/>

In this work we study the problem of efficiently isolating witnesses for the complexity classes NL and LogCFL, which are two well-studied complexity classes contained in P. We prove that if there is a L/poly randomized procedure with success probability at least 2/3 for isolating an s - t path in a given directed graph with a source sink pair (s, t) , then NL is contained in L/poly. By isolating a path we mean outputting a new graph on the same set of nodes such that exactly one s - t path from the original graph survives. Such an isolating procedure will naturally imply a UL/poly algorithm for reachability, but we prove that in fact this implies an L/poly algorithm. We also prove a similar result for the class LogCFL.

3.15 Enumerator polynomials: Completeness and Intermediate Complexity

Meena Mahajan (The Institute of Mathematical Sciences – Chennai, IN)

License  Creative Commons BY 3.0 Unported license
© Meena Mahajan

Joint work of Nitin Saurabh

Main reference M. Mahajan, N. Saurabh, “Some Complete and Intermediate Polynomials in Algebraic Complexity Theory”, ECCC TR16-38, 2016.

URL <http://eccc.hpi-web.de/report/2016/038/>

VNP, VP, VBP are central complexity classes in algebraic complexity theory. The notions of reductions and completeness are central to understanding the relationships between them. This talk will describe

1. polynomial families based on graph homomorphisms and complete for each of these classes,
2. polynomial families based on basic combinatorial NP-complete problems, and unless PH collapses, provably *intermediate* in nature,
3. a lower bound showing that to express the clique polynomial as a monotone projection of the permanent polynomial, exponential *blow-up* is required.

3.16 Toward the KRW Composition Conjecture: Cubic Formula Lower Bounds via Communication Complexity

Or Meir (University of Haifa, IL)

License © Creative Commons BY 3.0 Unported license
© Or Meir

Joint work of Or Meir, Irit Dinur

Main reference Irit Dinur, Or Meir, “Toward the KRW Composition Conjecture: Cubic Formula Lower Bounds via Communication Complexity”, in Proc. of the 31st Conf. on Computational Complexity (CCC 2016), LIPIcs, Vol. 50, pp. 3:1–3:51, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.

URL <http://dx.doi.org/10.4230/LIPIcs.CCC.2016.3>

One of the major challenges of the research in circuit complexity is proving super-polynomial lower bounds for de-Morgan formulas. Karchmer, Raz, and Wigderson suggested to approach this problem by proving that formula complexity behaves *as expected* with respect to the composition of functions. They showed that this conjecture, if proved, would imply super-polynomial formula lower bounds.

We prove a special case of this conjecture, in which one composes an arbitrary function with the parity function.

While this special case of the KRW conjecture was already proved implicitly in Hastad’s work on random restrictions, our proof seems more likely to be generalizable to other cases of the conjecture. In particular, our proof uses an entirely different approach, based on communication complexity technique of Karchmer and Wigderson.

3.17 Efficient Quantum Tomography and Longest Increasing Subsequences

Ryan O’Donnell (Carnegie Mellon University – Pittsburgh, US)

License © Creative Commons BY 3.0 Unported license
© Ryan O’Donnell

Joint work of Ryan O’Donnell, John Wright

In quantum mechanics, the state ρ of a d -dimensional particle is given by a $d \times d$ PSD matrix of trace 1. The *quantum tomography problem* is to estimate ρ accurately using as few *copies* of the state as possible. The special case of *quantum spectrum estimation* involves just estimating the eigenvalues $\alpha_1, \dots, \alpha_d$ of ρ , which form a probability distribution. By virtue of some representation theory, understanding these problems mostly boils down to understanding certain statistics of random words with i.i.d. letters drawn from the α_i distribution. These statistics involve longest increasing subsequences, and more generally, the shape of Young tableaux produced by the Robinson-Schensted-Knuth algorithm. In this talk we will discuss new probabilistic, combinatorial, and representation-theoretic tools for these problems, and the consequent new upper and lower bounds for quantum tomography.

3.18 Operator Scaling and Applications to Algebraic Complexity, Mathematics and Optimization

Rafael Oliveira (Princeton University, US)

License  Creative Commons BY 3.0 Unported license
© Rafael Oliveira

Joint work of Rafael Oliveira, Ankit Garg, Leonid Gurvits, Avi Wigderson

In this talk we shall explore the *non-commutative symbolic singularity problem*, and its myriad incarnations in commutative and non-commutative algebra, computational complexity, optimization and quantum information theory. We will describe an efficient algorithm solving all these related problems, and how its analysis combines ideas from all these areas. The problem these algorithms solve is non-convex, and we hope they will have many other applications.

3.19 On the Complexity of Generating Primes

Rahul Santhanam (University of Oxford, GB)

License  Creative Commons BY 3.0 Unported license
© Rahul Santhanam

Joint work of Igor Carboni Oliveira, Rahul Santhanam

The question of whether n -bit primes can be generated deterministically in time $\text{poly}(n)$ (or even $\text{subexp}(n)$) is a fundamental one in computational number theory. Despite much work on this problem, including the Polymath 4 project, no deterministic algorithm running in time better than $2^{n/2}$ is known.

We consider a relaxation of this question: *pseudo-deterministic* constructions, as defined and studied recently by Shafi Goldwasser and others. A zero-error pseudo-deterministic construction of primes in time $T(n)$ is a randomized algorithm, which for all large enough n , on input 1^n , halts in expected time $T(n)$ and outputs a *fixed* prime p_n (which does not depend on the randomness of the algorithm).

We show that either there is a deterministic sub-exponential time construction of infinitely many primes or there is a zero-error pseudo-deterministic construction of primes in sub-exponential time. In particular, this implies an unconditional zero-error pseudo-deterministic construction of infinitely many primes in sub-exponential time. The construction can be made deterministic under the assumption that $\text{ZPP}=\text{P}$, partially answering a question of the Polymath 4 project.

3.20 High rate locally-correctable and locally-testable codes with sub-polynomial query complexity

Shubhangi Saraf (Rutgers University – Piscataway, US)

License © Creative Commons BY 3.0 Unported license
© Shubhangi Saraf

Joint work of Shubhangi Saraf, Swastik Kopparty, Or Meir, Noga Ron-Zewi

Main reference S. Kopparty, O. Meir, N. Ron-Zewi, S. Saraf, “High rate locally-correctable and locally-testable codes with sub-polynomial query complexity”, ECCV TR15-68, 2016.

URL <http://eccv.hpi-web.de/report/2015/068/>

We study locally correctable and locally testable codes in the high rate regime. The tradeoff between the rate of a code and the locality/efficiency of its decoding and testing algorithms has been studied extensively in the past decade, with numerous applications to complexity theory and pseudorandomness.

In this talk I will discuss some recent results giving efficient sub-polynomial query decoding and testing algorithms for high rate error correcting codes. will also highlight some of the most interesting challenges that remain.

3.21 Algebraic independence over positive characteristic: New criterion and applications to locally low algebraic rank circuits

Nitin Saxena (Indian Institute of Technology – Kanpur, IN)

License © Creative Commons BY 3.0 Unported license
© Nitin Saxena

Joint work of Nitin Saxena, Anurag Pandey, Amit Sinhababu

Main reference A. Pandey, N. Saxena, A. Sinhababu, “Algebraic Independence over Positive Characteristic: New Criterion and Applications to Locally Low Algebraic Rank Circuits”, in Proc. of the 41st Int’l Symposium on Mathematical Foundations of Computer Science (MFCS 2016), LIPIcs, Vol. 58, pp. 74:1–74:15, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.

URL <http://dx.doi.org/10.4230/LIPIcs.MFCS.2016.74>

The motivation for this work comes from two problems – test algebraic independence of arithmetic circuits over a field of small characteristic, and generalize the structural property of algebraic dependence used by (Kumar, Saraf CCC’16) to arbitrary fields.

It is known that in the case of zero, or large characteristic, using a classical criterion based on the Jacobian, we get a randomized poly-time algorithm to test algebraic independence. Over small characteristic, the Jacobian criterion fails and there is no subexponential time algorithm known. This problem could well be conjectured to be in RP, but the current best algorithm puts it in $\text{NP}^{\#P}$ (Mittmann, Saxena, Scheiblechner Trans. AMS’14). Currently, even the case of two bivariate circuits over F_2 is open. We come up with a natural generalization of Jacobian criterion, that works over all characteristic. The new criterion is efficient if the underlying inseparable degree is promised to be a constant. This is a modest step towards the open question of fast independence testing, over finite fields, posed in (Dvir, Gabizon, Wigderson FOCS’07).

In a set of linearly dependent polynomials, any polynomial can be written as a linear combination of the polynomials forming a basis. The analogous property for algebraic dependence is false, but a property approximately in that spirit is named as *functional dependence* in (Kumar, Saraf CCC’16) and proved for zero or large characteristic. We show that functional dependence holds for arbitrary fields, thereby answering the open questions in (Kumar, Saraf CCC’16). Following them we use the functional dependence lemma to

prove the first exponential lower bound for locally low algebraic rank circuits for arbitrary fields (a model that strongly generalizes homogeneous depth-4 circuits). We also recover their quasipoly-time hitting-set for such models, for fields of characteristic smaller than the ones known before.

Our results show that approximate functional dependence is indeed a more fundamental concept than the Jacobian as it is field independent. We achieve the former by first picking a *good* transcendence basis, then translating the circuits by new variables, and finally approximating them by truncating higher degree monomials. We give a tight analysis of the *degree* of approximation needed in the criterion. To get the locally low algebraic rank circuit applications we follow the known shifted partial derivative based methods.

3.22 Pseudorandomness when the odds are against you

Ronen Shaltiel (University of Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Ronen Shaltiel

Joint work of Ronen Shaltiel, Sergei Artemenko, Russel Impagliazzo, Valentine Kabanets

Main reference S. Artemenko, R. Impagliazzo, V. Kabanets, R. Shaltiel, “Pseudorandomness when the odds are against you”, ECCC TR16-37, 2016.

URL <http://eccc.hpi-web.de/report/2016/037/>

A celebrated result by Impagliazzo and Wigderson is that under complexity theoretic hardness assumptions, every randomized algorithm can be transformed into one that uses only logarithmically many bits, with polynomial slowdown. Such algorithms can then be completely derandomized, with polynomial slowdown. In the talk I will discuss recent work attempting to extend this approach to:

1. Randomized algorithms that err with probability $1 - \epsilon$ for small ϵ . (Here, the goal is to minimize the number of random bits/slowdown as a function of ϵ).
2. Known SAT-solving randomized algorithms. (Here, polynomial slowdown is a deal breaker as it gives trivial algorithms that run in super exponential time).
3. Randomized algorithms that sample from probability distributions. (Here, the goal is to sample a statistically-close distribution using only few random bits).

3.23 Explicit two-source extractors for near-logarithmic min-entropy

Amnon Ta-Shma (Tel Aviv University, IL)

License  Creative Commons BY 3.0 Unported license
© Amnon Ta-Shma

Joint work of Amnon Ta-Shma, Avraham Ben-Aroya, Dean Doron

Main reference A. Ben-Aroya, D. Doron, A. Ta-Shma, “Explicit two-source extractors for near-logarithmic min-entropy”, ECCC TR16-88, 2016.

URL <http://eccc.hpi-web.de/report/2016/088/>

We explicitly construct extractors for two independent n -bit sources of $(\log n)^{1+o(1)}$ min-entropy. Previous constructions required either $\text{polylog}(n)$ min-entropy [CZ15] or five sources [Cohen16].

Our result extends the breakthrough result of Chattopadhyay and Zuckerman [CZ15] and uses the non-malleable extractor of Cohen [Cohen16]. The main new ingredient in our construction is a somewhere-random condenser with a small entropy gap, used as a sampler.

We construct such somewhere-random condensers using the error reduction mechanism of Raz et al. [RRV99] together with the high-error, constant degree dispersers of Zuckerman [Zuc06].

Using our framework and results Cohen and independently Li constructed 2-source extractors for even smaller min-entropies with the world record currently being $O(\log n \log \log n)$.

3.24 A one-sided error randomized protocol for Gap Hamming Distance problem

Nikolay K. Vereshchagin (NRU Higher School of Economics – Moscow, RU)

License  Creative Commons BY 3.0 Unported license
© Nikolay K. Vereshchagin

Assume that Alice has a binary string x and Bob a binary string y , both of length n . Their goal is to output 0, if x and y are at least L -close in Hamming distance, and output 1, if x and y are at least U -far in Hamming distance, where $L < U$ are some integer parameters known to both parties. If the Hamming distance between x and y lies in the interval (L, U) , they are allowed to output anything. This problem is called the Gap Hamming Distance (GHD). We study public-coin one-sided error communication complexity of this problem. The error with probability at most $1/2$ is allowed only for pairs at Hamming distance at least U . We establish the upper bound $O((L^2/U) \log L)$ and the lower bound $\Omega(L^2/U)$ for this complexity. These bounds differ only by a $O(\log L)$ factor.

The best upper bounds for communication complexity of GHD known before are the following. The upper bounds $O(L \log n)$ for one-sided error complexity and $O(L \log L)$ for two-sided error complexity, which do not depend on U and hold for all $U > L$. Our bound is better than these two bounds in the case when the ratio U/L is not bounded by a constant. The other known upper bound $O(L^2/(U-L)^2)$ holds for two-sided error complexity of GHD. If U is greater than $L + \sqrt{L}$ then this bound is better than ours, however it is for two-sided error. It is worth to note that all mentioned protocols run in one round.

From technical viewpoint, our achievement is a new protocol to prove that x, y are far on the basis of a large difference between distances from x and y to a randomly chosen string.

Our lower bound $\Omega(L^2/U)$ (for the one-sided error communication complexity of GHD) generalizes the lower bound $\Omega(U)$ for $U = O(L)$ known before.

3.25 Pointer chasing via triangular discrimination

Amir Yehudayoff (Technion – Haifa, IL)

License  Creative Commons BY 3.0 Unported license
© Amir Yehudayoff

Main reference A. Yehudayoff, “Pointer chasing via triangular discrimination”, ECCC TR16-151, 2016.
URL <http://eccc.hpi-web.de/report/2016/151/>

We prove an essentially sharp $\tilde{\Omega}(n/k)$ lower bound on the k -round distributional complexity of the k -step pointer chasing problem under the uniform distribution, when Bob speaks first. This is an improvement over Nisan and Wigderson’s $\tilde{\Omega}(n/k^2)$ lower bound. The proof is information theoretic, and a key part of it is using triangular discrimination instead of total variation distance; this idea may be useful elsewhere.

Participants

- Farid Ablayev
Kazan State University, RU
- Vikraman Arvind
The Institute of Mathematical Sciences, India, IN
- Markus Bläser
Universität des Saarlandes, DE
- Andrej Bogdanov
The Chinese University of Hong Kong, HK
- Arkadev Chattopadhyay
Tata Institute of Fundamental Research – Mumbai, IN
- Samir Datta
Chennai Mathematical Institute, IN
- Stephen A. Fenner
University of South Carolina – Columbia, US
- Michael A. Forbes
Stanford University, US
- Anna Gál
University of Texas – Austin, US
- Frederic Green
Clark University – Worcester, US
- Rohit Gurjar
Aalen University, DE
- Kristoffer Arnsfelt Hansen
Aarhus University, DK
- William Hoza
University of Texas – Austin, US
- Valentine Kabanets
Simon Fraser University – Burnaby, CA
- Marek Karpinski
Universität Bonn, DE
- Neeraj Kayal
Microsoft Research India – Bangalore, IN
- Pascal Koiran
ENS – Lyon, FR
- Swastik Kopparty
Rutgers University – Piscataway, US
- Arpita Korwar
University Paris-Diderot, FR
- Michal Koucký
Charles University – Prague, CZ
- Andreas Krebs
Universität Tübingen, DE
- Sophie Laplante
University Paris-Diderot, FR
- Nutan Limaye
Indian Institute of Technology – Mumbai, IN
- Meena Mahajan
The Institute of Mathematical Sciences, India, IN
- Pierre McKenzie
University of Montréal, CA
- Or Meir
University of Haifa, IL
- David A. Mix Barrington
University of Massachusetts – Amherst, US
- Ryan O’Donnell
Carnegie Mellon University – Pittsburgh, US
- Rafael Oliveira
Princeton University, US
- Chandan Saha
Indian Institute of Science – Bangalore, IN
- Rahul Santhanam
University of Oxford, GB
- Shubhangi Saraf
Rutgers University – Piscataway, US
- Nitin Saxena
Indian Institute of Technology – Kanpur, IN
- Uwe Schöning
Universität Ulm, DE
- Ronen Shaltiel
University of Haifa, IL
- Amnon Ta-Shma
Tel Aviv University, IL
- Thomas Thierauf
Hochschule Aalen, DE
- Jacobo Torán
Universität Ulm, DE
- Christopher Umans
CalTech – Pasadena, US
- Nikolay K. Vereshchagin
NRU Higher School of Economics – Moscow, RU
- Amir Yehudayoff
Technion – Haifa, IL
- Jeroen Zuiddam
CWI – Amsterdam, NL



Automated Algorithm Selection and Configuration

Edited by

Holger H. Hoos¹, Frank Neumann², and Heike Trautmann³

1 University of British Columbia, CA, hoos@cs.ubc.ca

2 University of Adelaide, AU, frank.neumann@adelaide.edu.au

3 Universität Münster, DE, trautmann@wi.uni-muenster.de

Abstract

This report documents the programme and the outcomes of Dagstuhl Seminar 16412 “Automated Algorithm Selection and Configuration”, which was held October 9–14, 2016 and attended by 34 experts from 10 countries. Research on automated algorithm selection and configuration has led to some of the most impressive successes within the broader area of empirical algorithmics, and has proven to be highly relevant to industrial applications. Specifically, high-performance algorithms for \mathcal{NP} -hard problems, such as propositional satisfiability (SAT) and mixed integer programming (MIP), are known to have a huge impact on sectors such as manufacturing, logistics, healthcare, finance, agriculture and energy systems, and algorithm selection and configuration techniques have been demonstrated to achieve substantial improvements in the performance of solvers for these problems. Apart from creating synergy through close interaction between the world’s leading groups in the area, the seminar pursued two major goals: to promote and develop deeper understanding of the behaviour of algorithm selection and configuration techniques and to lay the groundwork for further improving their efficacy. Towards these ends, the organisation team brought together a group of carefully chosen researchers with strong expertise in computer science, statistics, mathematics, economics and engineering; a particular emphasis was placed on bringing together theorists, empiricists and experts from various application areas, with the goal of closing the gap between theory and practice.

Seminar October 9–14, 2016 – <http://www.dagstuhl.de/16412>

1998 ACM Subject Classification I.2 Artificial Intelligence, I.2.2 Automatic Programming, I.2.6 Learning, I.2.8 Problem Solving, Control Methods, and Search, G.1.6 Optimization

Keywords and phrases algorithm configuration, algorithm selection, features, machine learning, optimisation, performance prediction

Digital Object Identifier 10.4230/DagRep.6.10.33

Edited in cooperation with Marius Lindauer

1 Executive Summary

Holger H. Hoos

Frank Neumann

Heike Trautmann

License  Creative Commons BY 3.0 Unported license
© Holger H. Hoos, Frank Neumann, and Heike Trautmann

The importance of high-performance algorithms, in particular for solving \mathcal{NP} -hard optimisation and decision problems, cannot be underestimated. Achievements in this area have substantial impact in sectors such as manufacturing, logistics, healthcare, finance, agriculture and energy systems – all of strategic importance to modern societies.



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

Automated Algorithm Selection and Configuration, *Dagstuhl Reports*, Vol. 6, Issue 10, pp. 33–74

Editors: Holger H. Hoos, Frank Neumann, and Heike Trautmann



Dagstuhl Reports

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

The development of effective automated algorithm selection and configuration techniques has been one of the major success stories in the area of empirical algorithmics in recent years. Building on a wide range of algorithmic approaches for problems such as propositional satisfiability (SAT) and mixed integer programming (MIP), these methods permit the selection of appropriate algorithms based on efficiently computable characteristic of a problem instance to be solved (algorithm selection) and the automatic determination of performance optimising parameter settings (algorithm configuration). In both cases, statistical models that enable performance predictions for previously unseen problem instances or parameter settings play a key enabling role; additionally, these models have other important uses, e.g., in load scheduling and distribution on large computer clusters.

The reach of those methods is illustrated by the fact that they have defined the state of the art in solving SAT, arguably the most prominent \mathcal{NP} -complete decision problem, for a decade (as witnessed by the results from the international SAT solver competitions (<http://www.satcompetition.org>), and more recently have been demonstrated to have the potential to achieve significant improvements over the long-standing state of the art in solving the TSP, one of the most widely studied \mathcal{NP} -hard optimisation problems [1]. Further very encouraging results have been achieved in recent years for continuous optimisation, AI planning and mixed integer programming problems.

The goal of the seminar was to foster research on algorithm selection and configuration, as well as on the underlying performance prediction methods, by bringing together researchers from the areas of artificial intelligence, theoretical computer science and machine learning in order to extend current studies to a much broader class of problems and build up the theoretical foundations of this important research area. On the foundational side, the seminar aimed at bridging the gap between experiments and theory in feature-based algorithm (runtime) analysis. In particular, we began investigating how mathematical and theoretical analyses can contribute to the experimentally driven research area of algorithm selection and configuration. We expect that studies following this initial exploration will bring together two of the currently most successful approaches for analysing heuristic search algorithms and ultimately achieve substantial impact in academic and industrial applications of algorithm configuration and selection techniques. Furthermore, we placed an emphasis on investigating automated algorithm selection and configuration approaches for multiobjective optimisation problems – an important, but largely unexplored area of investigation.

Background and Challenges: Algorithm Selection and Configuration for Combinatorial Problems

The design of algorithms for combinatorial optimisation and decision problems plays a key role in theoretical computer science as well as in applied algorithmics. These problems are frequently tackled using heuristic methods that perform extremely well on different classes of benchmark instances but usually do not have rigorous performance guarantees. Algorithm selection and configuration techniques have been applied to some of the most prominent \mathcal{NP} -hard combinatorial optimisation and decision problems, such as propositional satisfiability (SAT) and the travelling salesman problem (TSP).

Algorithm selection for SAT has first been explored in the seminal work on SATzilla [2, 3, 4], which was initially based on linear and ridge regression methods for performance prediction, but later moved to more sophisticated models based on cost-sensitive random forest classification [5]. Other successful methods use clustering techniques to identify the algorithm to be run on a given instance [6, 7]. As clearly evident from the results of SAT competitions, which are regularly held to assess and document the state of the art in SAT

solving, automated algorithm selection procedures effectively leverage the complementary strengths of different high-performance solvers and thus achieve substantial improvements over the best individual solvers [5].

As heuristic search algorithms often have numerous parameters that influence their performance, one of the classical questions is how to set parameters to optimise performance on a given class of instances. This per-set algorithm configuration problems can be solved using stochastic local search and model-based optimisation techniques [8, 9, 10], as well as racing techniques [11, 12], and these configuration methods have been demonstrated to yield substantial performance improvements to state-of-the-art algorithms for SAT, TSP, MIP, AI planning and several other problems (see, e.g., [13, 14, 15, 16]). Algorithm configuration techniques are now routinely used for optimising the empirical performance of solvers for a wide range of problems in artificial intelligence, operations research and many application areas (see, e.g., [17, 18]).

Initial work on combining algorithm selection and configuration techniques has shown significant promise [19, 20]; such combinations allow configuring algorithms on a per-instance basis [6, 7] and configuring algorithm selection methods (which themselves make use of many heuristic design choices) [21]. However, we see much room for further work along these lines. Other challenges concern the automated selection and configuration of mechanisms that adapt parameter settings while an algorithm is running and the configuration of algorithms for optimised scaling behaviour. Finally, a better theoretical foundation of algorithm selection and configuration approaches is desired and necessary. Initial steps into this direction were an important goal of this Dagstuhl seminar. In the following, we motivate and outline some of the challenges addressed in the course of the seminar.

Background and Challenges: Algorithm Selection for Continuous Black-Box Optimisation

Black-box function optimisation is a basic, yet intensely studied model for general optimisation tasks, where all optimisation parameters are real-valued. Work in this area has important practical applications in parameter and design optimisation and has also inspired some of the most successful general-purpose algorithm configuration techniques currently available [9].

Despite many years of research in metaheuristics, especially evolutionary algorithms, aimed at optimising black-box functions effectively, it is currently hardly possible to automatically determine a good optimisation algorithm for a given black-box function, even if some of its features are known. In single-objective (SO) black-box optimisation, it is therefore of considerable interest to derive rules for determining how problem properties influence algorithm performance as well as for grouping test problems into classes for which similar performance of the optimisation algorithms can be observed. Recent benchmarking experiments [22, 23] provide at best high-level guidelines for choosing a suitable algorithm type based on basic features that are known a priori, such as the number of dimensions of the given problem. However, the preference rules for algorithm selection thus obtained are very imprecise, and even for slight algorithm or problem variations, the resulting performance-induced ordering of different algorithms can change dramatically.

Exploratory Landscape Analysis (ELA, [24]) aims at improving this situation by deriving cheaply computable problem features based on which models relating features to algorithm performance can be constructed using benchmark experiments. The final goal is an accurate prediction of the best suited algorithm for an arbitrary optimisation problem based on the computed features. The concept is not entirely new; however, earlier approaches, such as fitness distance correlation (FDC) [25], have not been completely convincing.

A first idea to employ high-level (human expert designed) features, such as separability and modality, to characterize optimisation problems in an ELA context [26] was therefore refined by also integrating low-level features – e.g., based on convexity or the behaviour of local search procedures[27]. These effectively computable low-level features can be chosen from a wide range of easy to measure statistical properties. Suitably determined combinations of such features are expected to provide sufficient information to enable successful algorithm selection. Following recent results [27], this process is not necessarily costly in terms of function evaluations required for feature computation.

Additional, conceptually similar features were introduced in [28, 29, 30, 31]. In [32], a representative portfolio of four optimisation algorithms was constructed from the complete list of BBOB 2009/2010 candidates. Based on the low-level features a sufficiently accurate prediction of the best suited algorithm within the portfolio for each function was achieved. Recently, the feature set was extended based on the cell mapping concept in [33] by which a finite subdivision of the domain in terms of hypercubes is constructed. Most recently, the ELA approach, extended by several specific features, was successfully used to experimentally detect funnel structured landscapes in unknown black-box optimisation problems [34]. As it can be assumed that this information can be efficiently exploited to speed up the optimisation process, we expect ELA to contribute importantly to automated algorithm selection in single-objective black-box optimisation. (See [35] for a survey of related work.)

One major challenge in this area is the construction of a suitable algorithm portfolio together with an algorithm selection mechanism for unknown instances that generalises well to practical applications. For this purpose, suitable benchmark sets have to be derived and the costs of feature computations have to be kept as small as possible. Furthermore, theoretical foundation of the approaches is desired and necessary. The seminar aimed to make first steps in this direction.

Special focus: Algorithm selection for multiobjective optimisation

Some of the most challenging real-world problems involve the systematic and simultaneous optimisation of multiple conflicting objective functions – for example, maximising product quality and manufacturing efficiency, while minimising production time and material waste. To solve such problems, a large number of multiobjective optimisation (MOO) algorithms has been reported. Like single-objective (SO) algorithms, new MOO algorithms are claimed to outperform others by comparing the results over a limited set of test problems. Knowles et al. [36] started working on systematically deriving performance measures for EMOA and evaluating EMOA performance. Mersmann et al. [37] recently derived a systematic benchmarking framework according to similar work of [38] on benchmarking classification algorithms.

However, it is unlikely that any algorithm would outperform all others on a broader set of problems, and it is possible that the algorithm fails miserably on some of them. These results go usually unreported, leaving the algorithm's limitations unknown. This knowledge is crucial to avoid deployment disasters, gain theoretical insights to improve algorithm design, and ensure that algorithm performance is robustly described. Therefore, we see much value in the development of an algorithm selection and configuration framework for multiobjective optimisation. Successfully selecting the proper optimization algorithm for a multi-objective problem depends on detecting different problem characteristics, one of which is the multimodality of the induced landscape. In recent work [39], formal definitions were introduced for multimodality in multi-objective optimization problems in order to generalize the ELA framework to multi-objective optimization.

Significant progress has been made on single-objective (SO) problems of combinatorial and continuous nature as discussed above. However, these ideas are yet to be applied to the important class of MOO problems. We see five major avenues of exploration: (1) analysis on what makes MOO problems difficult; (2) design of features to numerically characterize MOO problems; (3) identification and visualization of strengths and weaknesses of state-of-the-art MOO algorithms; (4) methodology to assist the algorithm selection and configuration on (possibly expensive) real-world problems; (5) methodology to assist the design of tailored algorithms for real-world problems. An important aim of the seminar was to facilitate discussion of these directions.

Seminar structure and outcomes

The seminar was structured to balance short invited presentations with group breakout sessions and a generous amount of time set aside for informal discussions and spontaneously organised working groups at a ratio of about 2:1:1. Based on feedback obtained during and after the event, this structure worked well in fostering a vibrant atmosphere of intense and fruitful exchange and discussion. Presenters very successfully introduced important ideas, outlined recent results and open challenges, and facilitated lively discussion that provided much additional value. The afternoon group breakout sessions were particularly effective in addressing the challenges previously outlined as well as additional topics of interest that emerged during the seminar – thanks to the preparation and moderation by the session organisers as well as the lively participation of the attendees.

While it would be unreasonable to expect to exhaustively or conclusively address the substantial research challenges that inspired us to organise this Dagstuhl seminar, we believe that very significant progress has been achieved. As importantly, we feel that through this week-long event, an invaluable sharing of perspective and ideas has taken place, whose beneficial effects on the algorithm selection and configuration community and its work we hope to be felt for years to come. The following presentation abstracts and session summaries provided by the participants reflect the richness and depth of the scientific exchange facilitated by the seminar.

As organisers, we very much enjoyed working with presenters and session organisers, who greatly contributed to the success of the seminar, as did everyone who participated. Our thanks also go to the local team at Schloss Dagstuhl, who provided outstanding organisational support and a uniquely inspiring environment.

References

- 1 L. Kotthoff, P. Kerschke, H. Hoos, and H. Trautmann. Improving the state of the art in inexact TSP solving using per-instance algorithm selection. In C. Dhaenens, L. Jourdan, and M.-E. Marmion, editors, *Learning and Intelligent Optimization, 9th International Conference*, pages 202–217, Cham, 2015. Springer International Publishing. Publication status: Published.
- 2 E. Nudelman, K. Leyton-Brown, H. H Hoos, A. Devkar, and Y. Shoham. Understanding random SAT: Beyond the clauses-to-variables ratio. In *Principles and Practice of Constraint Programming–CP 2004*, pages 438–452. Springer Berlin Heidelberg, 2004.
- 3 E. Nudelman, K. Leyton-Brown, A. Devkar, Y. Shoham, and H. Hoos. Satzilla: An algorithm portfolio for SAT. *Solver description, SAT competition*, 2004, 2004.
- 4 L. Xu, F. Hutter, H. Hoos, and K. Leyton-Brown. SATzilla: Portfolio-based algorithm selection for SAT. *Journal of Artificial Intelligence Research*, 32:565–606, 2008.

- 5 L. Xu, F. Hutter, H.H. Hoos, and K. Leyton-Brown. Evaluating component solver contributions to portfolio-based algorithm selectors. In *Theory and Applications of Satisfiability Testing–SAT 2012*, pages 228–241. Springer Berlin Heidelberg, 2012.
- 6 S. Kadioglu, Y. Malitsky, M. Sellmann, and K. Tierney. Isac-instance-specific algorithm configuration. *ECAI*, 215:751–756, 2010.
- 7 Y. Malitsky, A. Sabharwal, H. Samulowitz, and M. Sellmann. Algorithm portfolios based on cost-sensitive hierarchical clustering. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 608–614. AAAI Press, 2013.
- 8 F. Hutter, H.H. Hoos, K. Leyton-Brown, and K.P. Murphy. An experimental investigation of model-based parameter optimisation: Spo and beyond. In *GECCO’09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 271–278, New York, NY, USA, 2009. ACM.
- 9 F. Hutter, H.H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In *Learning and Intelligent Optimization*, pages 507–523. Springer Berlin Heidelberg, 2011.
- 10 C. Ansótegui, M. Sellmann, and K. Tierney. A gender-based genetic algorithm for the automatic configuration of algorithms. *Principles and Practice of Constraint Programming–CP 2009*, pages 142–157, 2009.
- 11 M. Birattari, T. Stützle, L. Paquete, and K. Varrentapp. A racing algorithm for configuring metaheuristics. In *GECCO ’02: Proc. of the Genetic and Evolutionary Computation Conference*, pages 11–18, 2002.
- 12 M. Birattari, Z. Yuan, P. Balaprakash, and T. Stützle. F-race and iterated F-race: An overview. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, editors, *Empirical Methods for the Analysis of Optimization Algorithms*. Springer, 2010.
- 13 F. Hutter, M.T. Lindauer, A. Balint, S. Bayless, H.H. Hoos, and K. Leyton-Brown. The configurable SAT solver challenge (CSSC). *Artificial Intelligence*, Accepted for publication., 2015.
- 14 J. Styles and H. Hoos. Ordered racing protocols for automatically configuring algorithms for scaling performance. In *Proceedings of the 15th Conference on Genetic and Evolutionary Computation (GECCO-13)*, pages 551–558. ACM, 2013.
- 15 F. Hutter, H.H. Hoos, and K. Leyton-Brown. Automated configuration of mixed integer programming solvers. In *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, pages 186–202. Springer Berlin Heidelberg, 2010.
- 16 M. Vallati, C. Fawcett, A.E. Gerevini, H.H. Hoos, and A. Saetti. Automatic generation of efficient domain-optimized planners from generic parametrized planners. In *Sixth Annual Symposium on Combinatorial Search (SoCS-13)*, pages 184–192, 2013.
- 17 P. Lengauer and H. Mössenböck. The taming of the shrew: Increasing performance by automatic parameter tuning for java garbage collectors. In *Proceedings of the 5th ACM/SPEC International Conference on Performance Engineering, ICPE ’14*, pages 111–122, New York, NY, USA, 2014. ACM.
- 18 J.P. Dickerson, A.D. Procaccia, and T. Sandholm. Dynamic matching via weighted myopia with application to kidney exchange. In *Proceedings of the 28th National Conference on Artificial Intelligence (AAAI-14)*, pages 1340–1346, 2012.
- 19 L. Xu, H.H. Hoos, and K. Leyton-Brown. Hydra: Automatically configuring algorithms for portfolio-based selection. *Proceedings of the 24th Conference on Artificial Intelligence (AAAI-10)*, 10:210–216, 2010.
- 20 L. Xu, F. Hutter, H.H. Hoos, and K. Leyton-Brown. Hydra-MIP: Automated algorithm configuration and selection for mixed integer programming. *RCRA workshop on experimental evaluation of algorithms for solving problems with combinatorial explosion at the international joint conference on artificial intelligence (IJCAI)*, pages 16–30, 2011.

- 21 M. Lindauer, H. Hoos, F. Hutter and T. Schaub: AutoFolio: An Automatically Configured Algorithm Selector. *J. Artif. Intell. Res. (JAIR)* 53:745-778, 2015.
- 22 N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2009: Experimental setup. Technical Report RR-6828, INRIA, 2009.
- 23 N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.
- 24 O. Mersmann, B. Bischl, H. Trautmann, M. Preuss, C. Weihs, and G. Rudolph. Exploratory landscape analysis. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, GECCO '11, pages 829–836, New York, NY, USA, 2011. ACM.
- 25 T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.
- 26 M. Preuss and Th. Bartz-Beielstein. Experimental analysis of optimization algorithms: Tuning and beyond. In Y. Borenstein and A. Moraglio, editors, *Theory and Principled Methods for Designing Metaheuristics*. Springer, 2011.
- 27 O. Mersmann, M. Preuss, H. Trautmann, B. Bischl, and C. Weihs. Analyzing the BBOB results by means of benchmarking concepts. *Evolutionary Computation Journal*, 23(1):161–185, 2015.
- 28 M. A. Muñoz, M. Kirley, and S. K. Halgamuge. A meta-learning prediction model of algorithm performance for continuous optimization problems. In C. A. Coello Coello et al., editors, *Parallel Problem Solving from Nature – PPSN XII*, volume 7491 of *Lecture Notes in Computer Science*, pages 226–235. Springer, 2012.
- 29 T. Abell, Y. Malitsky, and K. Tierney. Features for exploiting black-box optimization problem structure. In Giuseppe Nicosia and Panos Pardalos, editors, *Learning and Intelligent Optimization*, Lecture Notes in Computer Science, pages 30–36. Springer, 2013.
- 30 R. Morgan and M. Gallagher. Using landscape topology to compare continuous metaheuristics: A framework and case study on EDAs and ridge structure. *Evolutionary Computation*, 20(2):277–299, 2012.
- 31 M.A. Munoz, M. Kirley, and S.K. Halgamuge. Exploratory landscape analysis of continuous space optimization problems using information content. *Evolutionary Computation, IEEE Transactions on*, 19(1):74–87, Feb 2015.
- 32 B. Bischl, O. Mersmann, H. Trautmann, and M. Preuss. Algorithm selection based on exploratory landscape analysis and cost-sensitive learning. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, GECCO '12, pages 313–320. ACM, 2012.
- 33 P. Kerschke, M. Preuss, C. Hernández, O. Schütze, J. Sun, C. Grimme, G. Rudolph, B. Bischl, and H. Trautmann. Cell mapping techniques for exploratory landscape analysis. In A Tantar et al., editors, *EVOLVE — A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V*, volume 288 of *Advances in Intelligent Systems and Computing*, pages 115–131. Springer, 2014.
- 34 P. Kerschke, M. Preuss, S. Wessing, and H. Trautmann. Detecting funnel structures by means of exploratory landscape analysis. In *Proceedings of the*, pages 265–272, New York, NY, USA, 2015. ACM. Publication status: Published.
- 35 M. A. Muñoz, Y. Sun, M. Kirley, and S. K. Halgamuge. Algorithm selection for black-box continuous optimization problems: A survey on methods and challenges. *Information Sciences*, 317:224 – 245, 2015.
- 36 J. Knowles, L. Thiele, and E. Zitzler. A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers. TIK Report 214, Computer Engineering and Networks Laboratory, ETH Zurich, 2006.

- 37 O. Mersmann, H. Trautmann, B. Naujoks, and C. Weihs. Benchmarking evolutionary multiobjective optimization algorithms. In *IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE, 2010.
- 38 K. Hornik and D. Meyer. Deriving consensus rankings from benchmarking experiments. In R. Decker and H.-J. Lenz, editors, *Advances in Data Analysis (Proc. of the 30th Ann. Conf. of the Gesellschaft für Klassifikation)*, pages 163–170. Springer, Berlin, 2007.
- 39 P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. Deutz, H. Trautmann and M. Emmerich. Towards Analyzing Multimodality of Multiobjective Landscapes. In *Proceedings of the 14th International Conference on Parallel Problem Solving from Nature (PPSN XIV)*, pages 962–972. Lecture Notes in Computer Science, Springer, 2016.

2 Table of Contents

Executive Summary

Holger H. Hoos, Frank Neumann, and Heike Trautmann 33

Presentations

Continuous Black-Box Optimization: How Large Are The GAPS?
Anne Auger 44

Tuning using Multiple Criteria – Forwarding more Information to the Configurator
Aymeric Blot 44

Some Ideas on Industrial Applications of Automated Optimizer Design
Thomas Bäck 44

Learning the Right Mutation Strength on-the-Fly
Benjamin Doerr 45

Self-Adjusting Choice of Population Size and Mutation Strengths in Discrete Optimization
Carola Doerr 46

Spotlight on Rumor Spreading
Carola Doerr 46

Algorithm Subset Selection as a Portfolio Investment Problem
Michael Emmerich 46

Optimisation Algorithm Design: A Control Engineering Perspective
Carlos M. Fonseca 47

Models of Large Real-world Networks
Tobias Friedrich 48

Needed: Hard and Killer problems
Marcus Gallagher 48

Fixed-Parameter Single Objective Search Heuristics for Minimum Vertex Cover
Wanru Gao 48

AutoML – with a focus on deep learning
Frank Hutter 49

Applications in Hospital
Laetitia Jourdan 49

Extending Exploratory Landscape Analysis Toward multiobjective and Multimodal Problems
Pascal Kerschke and Mike Preuss 50

Deep Parameter Configuration (joint work with UCL)
Lars Kotthoff 50

Algorithm Portfolios: Four Key Questions
Kevin Leyton-Brown 51

Multiobjective Optimization from the Perspective of Game Theory
Kevin Leyton-Brown 51

| | |
|---|----|
| Combining Algorithm Selection and Configuration: Per-Instance Algorithm Configuration <i>Marius Lindauer</i> | 51 |
| Network on Selection and Configuration: COSEAL <i>Marius Lindauer</i> | 52 |
| Challenges in Automated Algorithm Design: Representativeness, one-shot expensive scenarios, parameter importance and sensitivity, and human-in-the-loop <i>Manuel López-Ibáñez</i> | 53 |
| Building and exploiting a non-parametric problem space <i>Andres Munoz Acosta</i> | 53 |
| Automated Selection of Tree Decompositions <i>Nysret Musliu</i> | 54 |
| Feature-Based Diversity Optimization for Problem Instance Classification <i>Samadhi Nethmini Nallaperuma</i> | 54 |
| Algorithm Selection in Music Data Analysis <i>Günter Rudolph</i> | 55 |
| Cognitive Assistant for Data Scientist <i>Horst Samulowitz</i> | 55 |
| From off-line to on-line feature-based parameter tuning <i>Marc Schoenauer</i> | 56 |
| Cognitive Computing <i>Meinolf Sellmann</i> | 56 |
| Automated generation of high-performance heuristics from flexible algorithm frameworks: Challenges and Perspectives <i>Thomas Stützle</i> | 57 |
| Bringing the human back in the loop <i>Joaquin Vanschoren</i> | 57 |
| A Generic Bet-and-run Strategy for Speeding Up Traveling Salesperson and Minimum Vertex Cover <i>Markus Wagner</i> | 58 |
| Reducing the size of large instance repositories <i>Markus Wagner</i> | 58 |
| Accelerating Algorithm Development <i>Simon Wessing</i> | 59 |

Breakout Sessions and Working Groups

| | |
|---|----|
| All you ever wanted to know/ask a theoretician and All you ever wanted to know/ask a practitioner <i>Anne Auger and Carola Doerr</i> | 59 |
| Controlled Problem Instance Generation <i>Marcus Gallagher</i> | 60 |
| Describing / Characterizing Landscapes of multiobjective Optimization Problems <i>Pascal Kerschke</i> | 61 |

| | |
|---|----|
| Portfolio-based Methods for Algorithm Selection <i>Kevin Leyton-Brown</i> | 63 |
| What can we learn from algorithm selection data? <i>Marius Lindauer and Lars Kotthoff</i> | 64 |
| (ELA features for) multimodal optimization <i>Mike Preuß</i> | 65 |
| Online and Adaptive Methods <i>Marc Schoenauer</i> | 66 |
| Real-world Applications of Meta-Algorithmics <i>Meinolf Sellmann</i> | 69 |
| Pitfalls and Best Practices for Algorithm Configuration <i>Marius Lindauer and Frank Hutter</i> | 70 |
| Multiobjective Optimisation Algorithm Selection and Configuration <i>Carlos M. Fonseca and Manuel López-Ibáñez</i> | 72 |
| Participants | 74 |

3 Presentations

3.1 Continuous Black-Box Optimization: How Large Are The GAPS?

Anne Auger (INRIA Saclay – Orsay, FR)

License  Creative Commons BY 3.0 Unported license
© Anne Auger

This talk was motivated by the breakout session on Theory versus Practice or “All you ever wanted to know/ask a theoretician and All you ever wanted to know/ask a practitioner” where the observation was made that gaps between theory and practice is larger or smaller depending on the domain and community.

We have discussed for the domain of continuous black-box optimization or adaptive stochastic search algorithms where theory stands with respect to practice. In particular we have sketched how theoretical results on linear convergence relate to practice and how they are motivated by practice. We have also highlighted a few lessons from theory to practice. Last we have discussed this gap between communities tackling the same problem namely the Derivative Free Optimization community and the Evolutionary Computation community and how there are signs that this gap in becoming narrower and narrower.

3.2 Tuning using Multiple Criteria – Forwarding more Information to the Configurator

Aymeric Blot (INRIA Lille, FR)

License  Creative Commons BY 3.0 Unported license
© Aymeric Blot

Joint work of Aymeric Blot, Holger Hoos, Marie-Éléonore Kessaci-Marmion, Laetitia Jourdan, Heike Trautmann

In automatic algorithm configuration, a single performance indicator of the target algorithm is usually forwarded to the configurator. We discuss problems and possible solutions in cases where more than a single indicator might be needed. The highlighted solution is MO-ParamILS, a configurator specifically designed for the purpose of performing the configuration process using Pareto dominance on multiple performance indicators.

3.3 Some Ideas on Industrial Applications of Automated Optimizer Design

Thomas Bäck (Leiden University, NL)

License  Creative Commons BY 3.0 Unported license
© Thomas Bäck

Joint work of Sander van Rijn, Hao Wang, Matthijs van Leeuwen, Thomas Bäck
Main reference S. van Rijn, H. Wang, M. van Leeuwen, T. Bäck, “Evolving the Structure of Evolution Strategies”, arXiv:1610.05231v1 [cs.NE], 2016.
URL <https://arxiv.org/abs/1610.05231v1>

Many industrial applications are represented by simulation models which require enormous computational effort for computing a single objective function value. Motivated by such applications, e.g., in the automotive industry, an important requirement for optimization

algorithms can be to deliver large improvements with the smallest number of function evaluations possible.

From an automated optimizer design perspective, I discuss some preliminary experiments on the automatic configuration of algorithmic variants of modern evolutionary strategies. As these results illustrate, it is possible to significantly improve performance by configuring the components of evolutionary strategies optimally.

The talk then presents my vision on how to extend this approach towards automated optimizer design for simulation-based function classes. This might involve response surface modeling, exploratory feature analysis, machine learning, and aspects of genetic programming and grammatical evolution – plus likely a number of additional techniques for this challenging problem.

3.4 Learning the Right Mutation Strength on-the-Fly

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

License © Creative Commons BY 3.0 Unported license
© Benjamin Doerr

Joint work of Benjamin Doerr, Carola Doerr, Jing Yang

When using the classic standard bit mutation operator, parent and offspring differ in a random number of bits, distributed according to a binomial law. This has the advantage that all Hamming distances occur with some positive probability, hence this operator can be used, in principle, for all fitness landscapes. The downside of this “one-size-fits-all” approach, naturally, is a performance loss caused by the fact that often not the ideal number of bits is flipped. Still, the fear of getting stuck in local optima has made standard bit mutation become the preferred mutation operator.

In this work we show that a self-adjusting choice of the number of bits to be flipped can both avoid the performance loss of standard bit mutation and avoid the risk of getting stuck in local optima. We propose a simple mechanism to adaptively learn the currently optimal mutation strength from previous iterations. This aims both at exploiting that generally different problems may need different mutation strengths and that for a fixed problem different strengths may become optimal in different stages of the optimization process.

We experimentally show that our simple hill climber with this adaptive mutation strength outperforms both the randomized local search heuristic and the (1+1) evolutionary algorithm on the LeadingOnes function and on the minimum spanning tree problem. We show via mathematical means that our algorithm is able to detect precisely (apart from lower order terms) the complicated optimal fitness-dependent mutation strength recently discovered for the OneMax function. With its self-adjusting mutation strength it thus attains the same runtime (apart from $o(n)$ lower-order terms) and the same (asymptotic) 13% fitness-distance improvement over RLS that was recently obtained by manually computing the optimal fitness-dependent mutation strength.

This talk is based on joint work with Carola Doerr (Paris 6) and Jing Yang (École Polytechnique)

References

- 1 Benjamin Doerr, Carola Doerr, Jing Yang. k -Bit Mutation with Self-Adjusting k Outperforms Standard Bit Mutation. In *Parallel Problem Solving from Nature – PPSN XIV*, pages 824–834, 2016.

3.5 Self-Adjusting Choice of Population Size and Mutation Strengths in Discrete Optimization

Carola Doerr (CNRS and University Pierre & Marie Curie – Paris, FR)

License  Creative Commons BY 3.0 Unported license
© Carola Doerr

In most evolutionary algorithms there are a number of parameters to be chosen, e.g., the population size, the mutation strength, the crossover rate, etc. While it seems quite intuitive that different parameter choices can be optimal in the different stages of the optimization process, little theoretical evidence exist to support this claim for discrete optimization problems. In two recent works [Doerr/Doerr, Optimal Parameter Choices Through Self-Adjustment: Applying the 1/5-th Rule in Discrete Settings, GECCO 2015] and [Doerr/Doerr/Kötzing: Provably Optimal Self-adjusting Step Sizes for Multi-valued Decision Variables, PPSN 2016] we propose a simple success-based update rule for the population size and the mutation strength, respectively. In both these works we show that the self-adjusting parameter choice yields a better performance than any (!) static parameter choice. The update rule is inspired by the classical one-fifth rule from continuous optimization. Based on joint work with Benjamin Doerr (Ecole Polytechnique, France) and Timo Koetzing (HPI Potsdam, Germany)

3.6 Spotlight on Rumor Spreading

Carola Doerr (CNRS and University Pierre & Marie Curie – Paris, FR)

License  Creative Commons BY 3.0 Unported license
© Carola Doerr

In this short talk we briefly discuss the rumor spreading problem and the main objectives that we are after when designing algorithms for it.

Rumor spreading aims at distributing information in networks via so-called PUSH operations. Informed nodes are allowed to call others to inform them. We aim at protocols that are fast, need few calls, are robust with respect to node and edge crashes, and hopefully simple. Existing works analyze rumor spreading algorithms on different types of graphs, e.g., social networks and dynamically changing graphs.

3.7 Algorithm Subset Selection as a Portfolio Investment Problem

Michael Emmerich (Leiden University, NL)

License  Creative Commons BY 3.0 Unported license
© Michael Emmerich

Joint work of Michael Emmerich, Iryna Yevseyeva

The problem of optimization algorithm selection (and configuration) can be formulated in a similar way than a financial portfolio investment problem with risk and expected return. Depending on the practical setting in which the optimization algorithms are applied, different scenarios and problem formulations can be of interest.

Here, as a pars pro toto for a larger class of problem formulations, one particular scenario is discussed. The chosen example formulation is motivated by problems that occur in logistics planning, and in real-world environments of computer-aided product design, and it reads as

follows: Prior to performing an time-expensive optimization task, a single algorithm or a subset of algorithms has to be selected from a library or algorithm portfolio. The algorithm (or the subset of algorithms) are then submitted to a parallel computation cluster where they have to solve an a priori unknown problem instance. After a pre-assigned time the results achieved by all selected algorithms are collected and the best result is chosen.

This scenario leads to a theoretical problem formulation where each algorithm is viewed as an investment (asset) and its result, which has to be maximized, is viewed as the financial return of this investment. Due to the uncertainty about the particular problem instance that will have to be solved, the return will be considered as a random variable. Given a single algorithm the expected value of the return has to be maximized and the risk, which is related to the variance of the return, is to be minimized. For a subset of algorithms, the expected maximum of the return has to be maximized, and the risk related to this value needs to be minimized. The computation of the risk term requires covariances of the returns of the algorithms, noting that diversification will usually be beneficial for reducing the risk.

There is a rich theory on the solution of such multiobjective portfolio selection problems, which can be transferred to the algorithm selection domain. However, there are also challenges to be overcome, such as the elicitation or estimation of probability distributions and the approximation or computation of the efficient set when it comes to large algorithm libraries or instance spaces, as they would need to be considered in algorithm tuning or configuration.

3.8 Optimisation Algorithm Design: A Control Engineering Perspective

Carlos M. Fonseca (University of Coimbra, PT)

License  Creative Commons BY 3.0 Unported license
© Carlos M. Fonseca

Joint work of Cláudia R. Correa, Elizabeth F. Wanner, Carlos M. Fonseca, Rodrigo T. N. Cardoso, Ricardo H. C. Takahashi

Main reference C. R. Correa, E. F. Wanner, and C. M. Fonseca, “Lyapunov design of a simple step-size adaptation strategy based on success”, in Proc. of the 14th Int’l Conf. on Parallel Problem Solving from Nature – PPSN XIV, LNCS, Vol. 9921, pp. 101–110, Springer, 2016.

URL http://dx.doi.org/10.1007/978-3-319-45823-6_10

Currently, most practically-relevant metaheuristic algorithms, and Evolutionary Algorithms (EAs) in particular, are not amenable to analysis with the available theoretical tools. On the other hand, EA theory has focused largely on asymptotic and time-complexity results in ideal or much simplified scenarios, which are not immediately useful to practitioners.

An alternative route for theoretical development is to approach the design of such optimisation algorithms from a control engineering perspective, where determining algorithm parameters is *the* purpose of the analysis, and algorithms must be designed with analysis in mind. The fact that optimisation algorithms are inherently dynamical systems further substantiates this point of view. Moreover, theory should support the use of numerical methods to extend the analysis to larger and/or more complex scenarios before experimentation becomes the only practical alternative.

This perspective will be illustrated with the design of a simple step-size adaptation strategy based on success and failure events [1]. A Lyapunov synthesis procedure is used to obtain both a performance guarantee and tuned adaptation parameter values. The method relies on the numerical optimisation of an analytically-derived performance index.

Acknowledgement. This work was partially supported by national funds through the Portuguese Foundation for Science and Technology (FCT) and by the European Regional

Development Fund (FEDER) through COMPETE 2020 – Operational Program for Competitiveness and Internationalisation (POCI).

References

- 1 C. R. Correa, E. F. Wanner, C. M. Fonseca, “Lyapunov design of a simple step-size adaptation strategy based on success,” in Proc. of the 14th Int’l Conf. on Parallel Problem Solving from Nature – PPSN XIV, LNCS, Vol. 9921, pp. 101–110, Springer, 2016.

3.9 Models of Large Real-world Networks

Tobias Friedrich (Hasso-Plattner-Institut – Potsdam, DE)

License  Creative Commons BY 3.0 Unported license
© Tobias Friedrich

The node degrees of large real-world networks often follow a power-law distribution. Such scale-free networks can be social networks, internet topologies, the web graph, power grids, or many other networks from literally hundreds of domains. The talk introduced several mathematical models of scale-free networks (e.g. preferential attachment graphs, Chung-Lu graphs, hyperbolic random graphs), showed some of their properties (e.g. diameter, average distance, clustering), and discussed how these properties influence algorithm selection and configuration.

3.10 Needed: Hard and Killer problems

Marcus Gallagher (The University of Queensland – Brisbane, AU)

License  Creative Commons BY 3.0 Unported license
© Marcus Gallagher

In this short talk, I would like to raise issues around the nature of optimization problems. What sorts of benchmark problems are needed to extract maximum value from our experiments? Do we need hard problems (but with rich structure) to solve and where are they (especially in the continuous case)? Finally, what are the so-called “killer apps” for the field?

3.11 Fixed-Parameter Single Objective Search Heuristics for Minimum Vertex Cover

Wanru Gao (University of Adelaide, AU)

License  Creative Commons BY 3.0 Unported license
© Wanru Gao

Joint work of Wanru Gao, Tobias Friedrich, Frank Neumann

Main reference W. Gao, T. Friedrich, F. Neumann, “Fixed-Parameter Single Objective Search Heuristics for Minimum Vertex Cover”, in Proc. of the 14th Int’l Conf. on Parallel Problem Solving from Nature – PPSN XIV, LNCS, Vol. 9921, pp. 740–750, Springer, 2016.

URL http://dx.doi.org/10.1007/978-3-319-45823-6_69

We consider how well-known branching approaches for the classical minimum vertex cover problem can be turned into randomized initialization strategies with provable performance guarantees and investigate them by experimental investigations. Furthermore, we show

how these techniques can be built into local search components and analyze a basic local search variant that is similar to a state-of-the-art approach called NuMVC. Our experimental results for the two local search approaches show that making use of more complex branching strategies in the local search component can lead to better results on various benchmark graphs.

3.12 AutoML – with a focus on deep learning

Frank Hutter (Universität Freiburg, DE)

License © Creative Commons BY 3.0 Unported license
© Frank Hutter

The rapid growth of machine learning (ML) applications has created a demand for off-the-shelf ML methods that can be used easily and without expert knowledge. I first briefly review the successful approach of casting this problem as an optimization problem on top of a highly-parameterized ML framework. Then, I focus on possible extensions of this approach that could scale it up to achieve fully automated deep learning: reasoning across datasets, subsets of data, and initial time steps; online hyperparameter control; and automatically deriving insights. Many of these extensions have a direct correspondence in optimizing hard combinatorial problem solvers.

3.13 Applications in Hospital

Laetitia Jourdan (INRIA Lille, FR)

License © Creative Commons BY 3.0 Unported license
© Laetitia Jourdan

Joint work of Clarisse Dhaenens, Laetitia Jourdan

Main reference J. Jacques, J. Taillard, D. Delerue, C. Dhaenens, L. Jourdan, “Conception of a dominance-based multiobjective local search in the context of classification rule mining in large and imbalanced data sets”, *Applied Soft Computing*, Vol. 34, pp. 705–720, Elsevier, 2015.

URL <http://dx.doi.org/10.1016/j.asoc.2015.06.002>

Main reference K. Seridi, L. Jourdan, E.-G. Talbi, “Using multiobjective optimization for biclustering microarray data”, *Applied Soft Computing*, Vol. 33, pp. 239–249, 2015.

URL <http://dx.doi.org/10.1016/j.asoc.2015.03.060>

Main reference J. Hamon, J. Jacques, L. Jourdan, C. Dhaenens, “Knowledge Discovery in Bioinformatics”, in *Handbook of Computational Intelligence*, pp. 1211–1223, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-662-43505-2_61

Main reference J. Jacques, J. Taillard, D. Delerue, L. Jourdan, C. Dhaenens, “The benefits of using multi-objectivization for mining pittsburgh partial classification rules in imbalanced and discrete data”, in *Proc. of the 15th Annual Conf. on Genetic and Evolutionary Computation (GECCO 2013)*, pp. 543–550, ACM, 2013.

URL <https://dx.doi.org/10.1145/2463372.2463432>

Hospital offers a lot of real world problems for the optimization and machine learning community. A lot of classical operation research problems can be found but often with a lot of additional constraints, presence of uncertainty and even dynamism of data. Concerning machine learning problems, there are very specific like classification on imbalanced data, bi-clustering and often solvers are not available in classical framework or they cannot cope the dimension of the data. All these problems can be modelled as optimisation problems often multiobjective problems. As final output should be software for non-domain specialists, automated configuration is required BUT how to realized it when there is only one dataset available, dataset that is often available only inside the hospital. Additionally, the robustness

of the proposed solutions is very important, as it can be critical for the hospital, how practitioners can assess the sensitivity of the found algorithms ?

3.14 Extending Exploratory Landscape Analysis Toward multiobjective and Multimodal Problems

Pascal Kerschke (Universität Münster, DE) and Mike Preuß (Universität Münster, DE)

License © Creative Commons BY 3.0 Unported license
© Pascal Kerschke and Mike Preuss

Main reference O. Mersmann, B. Bischl, H. Trautmann, M. Preuß, C. Weihs, G. Rudolph, “Exploratory landscape analysis”, in Proc. of the 13th Annual Conf. on Genetic and Evolutionary Computation (GECCO 2011), pp. 829–836, ACM, 2011.

URL <https://dx.doi.org/10.1145/2001576.2001690>

Main reference P. Kerschke, M. Preuß, S. Wessing, H. Trautmann, “Low-Budget Exploratory Landscape Analysis on Multiple Peaks Models”, in Proc. of the 18th Annual Conf. on Genetic and Evolutionary Computation (GECCO 2016), pp. 229–236, ACM, 2016.

URL <https://dx.doi.org/10.1145/2908812.2908845>

Main reference P. Kerschke, H. Wang, M. Preuß, C. Grimme, A. H. Deutz, H. Trautmann, M. Emmerich, “Towards Analyzing Multimodality of Continuous Multiobjective Landscapes”, in Proc. of the 14th Int’l Conf. on Parallel Problem Solving from Nature – PPSN XIV, LNCS, Vol. 9921, pp. 962–972, Springer, 2016.

URL http://dx.doi.org/10.1007/978-3-319-45823-6_90

Selecting the best suited algorithm for an optimization problem is usually a complex and difficult task, especially for expensive Black-Box problems. The Exploratory Landscape Analysis (ELA) approach extracts – not necessarily intuitively understandable – landscape features based on a (usually rather small) initial sample of observations from the underlying optimization problem. In case of population based algorithms, a well distributed initial population may be used as sample data for computing these features, which may then be used to enhance the algorithm selection model. So far, ELA is mostly used in the context of continuous, single-objective, global optimization problems – but it shall be transferred also to other domains. Next to a minimal introduction and report on the current state of ELA, we highlight the possibilities to extend it onto multiobjective and multimodal optimization.

3.15 Deep Parameter Configuration (joint work with UCL)

Lars Kotthoff (University of British Columbia – Vancouver, CA)

License © Creative Commons BY 3.0 Unported license
© Lars Kotthoff

Algorithm configuration has been limited to parameters that the programmer intentionally exposes. Automatic algorithm configuration makes it feasible to efficiently explore ever larger parameter spaces, but the mindset of programmers is still that parameters should be exposed sparingly as they put additional burden on the user. We leverage techniques from software engineering to expose additional parameters from the source of an algorithm, thus increasing the potential gains for algorithm configuration.

3.16 Algorithm Portfolios: Four Key Questions

Kevin Leyton-Brown (University of British Columbia – Vancouver, CA)

License  Creative Commons BY 3.0 Unported license
© Kevin Leyton-Brown

This talk considered four key questions:

1. How useful is my solver? I argued this is well answered using the Shapley value.
2. How useful is my algorithm selector? I argued that one needs to be careful to avoid using a biased estimate of VBS performance.
3. How useful are my selector's features? I argued that just size-based features are often enough.
4. How useful is my benchmark? I argued that heterogeneity may produce artificially easy test data.

3.17 Multiobjective Optimization from the Perspective of Game Theory

Kevin Leyton-Brown (University of British Columbia – Vancouver, CA)

License  Creative Commons BY 3.0 Unported license
© Kevin Leyton-Brown

I described what multiobjective optimization means to a game theorist. I discussed von Neumann-Morgenstern utility theory, multiattribute utility, noncooperative game theory. I discussed the solution concepts Pareto optimality, stability concepts like Nash equilibrium, robustness concepts like maxmin, and minimax regret.

3.18 Combining Algorithm Selection and Configuration: Per-Instance Algorithm Configuration

Marius Lindauer (Universität Freiburg, DE)

License  Creative Commons BY 3.0 Unported license
© Marius Lindauer

Algorithm configuration and algorithm selection perform well in different use cases, namely homogeneous instance sets with large parameter configuration spaces vs heterogeneous instances with a small finite portfolio of algorithms. One way to combine algorithm selection and configuration is for example to apply configuration on top of selection [1]. However, to deal with heterogeneous instances (e.g., hard combinatorial problems, machine learning data sets or environmental variables) and an algorithm with a large parameter configuration space, we need a direct combination of configuration and selection:

Per-Instance algorithm configuration (PIAC) approaches (such as ISAC [2] and Hydra [3]) were proposed already some years ago. In the meantime, relevant techniques for PIAC made substantial progress, e.g., prediction of performance [4], quantification of homogeneity [5] and feature-parameters mappings [6]. Using recent advances in these areas, I believe that we can do much better than previous approaches in generating robust algorithms that adapt their parameter settings on a per-instance base.

References

- 1 Marius Lindauer, Holger H. Hoos, Frank Hutter, Torsten Schaub: *AutoFolio: An Automatically Configured Algorithm Selector*. J. Artif. Intell. Res. (JAIR) 53:745-778 (2015)
- 2 Carlos Ansotegui, Joel Gabas, Yuri Malitsky, Meinolf Sellmann: *MaxSAT by improved instance-specific algorithm configuration*. Artif. Intell. 235: 26-39 (2016)
- 3 Lin Xu, Holger Hoos, Kevin Leyton-Brown: *Hydra: Automatically Configuring Algorithms for Portfolio-Based Selection*. AAAI 2010
- 4 Frank Hutter, Lin Xu, Holger H. Hoos, Kevin Leyton-Brown: *Algorithm runtime prediction: Methods and evaluation*. Artif. Intell. 206: 79-111 (2014)
- 5 Marius Schneider, Holger H. Hoos: *Quantifying Homogeneity of Instance Sets for Algorithm Configuration*. LION 2012: 190-204
- 6 Jakob Bossek, Bernd Bischl, Tobias Wagner, Günter Rudolph: *Learning Feature-Parameter Mappings for Parameter Tuning via the Profile Expected Improvement*. GECCO 2015: 1319-1326

3.19 Network on Selection and Configuration: COSEAL

Marius Lindauer (Universität Freiburg, DE)

License  Creative Commons BY 3.0 Unported license
© Marius Lindauer

Since algorithm selection and algorithm configuration is widely applicable in many domains (including e.g., machine learning, hard combinatorial problems and continuous optimization), there are sub-communities in all these domains that use automatic selection and configuration of algorithms to improve the performance of their algorithms. Unfortunately, these communities were not well connected, even though they worked on similar problems. To change this, the COSEAL group (COnfiguration and SElection of ALgorithms)¹ was founded three years ago to create a research network on automatic selection and configuration of any kind of algorithm.

To encourage exchange of progress and expert knowledge between the different communities, the COSEAL group has an active mailing list and an annual workshop meeting. The mailing list is an open platform where everyone can join. Its intended use includes the announcement of new important results, tools, and to request help for newcomers. Similar to our Dagstuhl seminar, the goal of the workshops is less to promote newly published papers but to discuss on-going projects and open questions. To this end, the workshop includes sessions for presentations, posters and discussions. One of the successful projects of the COSEAL group was also launched at the first COSEAL workshop: the creation of a benchmark library for algorithm selection, called ASlib [1]. Furthermore, the COSEAL website offers overviews and literature links for algorithm selection and configuration for new researchers.

References

- 1 Bischl, B., Kerschke, P., Kotthoff, L., Lindauer, M., Malitsky, Y., Frechétte, A., Hoos, H., Hutter, F., Leyton-Brown, K., Tierney, K. and Vanschoren, J. ASlib: *A Benchmark Library for Algorithm Selection* In: Artificial Intelligence Journal (AIJ) 237 (2016): 41-58

¹ <http://www.coseal.net>

3.20 Challenges in Automated Algorithm Design: Representativeness, one-shot expensive scenarios, parameter importance and sensitivity, and human-in-the-loop

Manuel López-Ibáñez (*Univ. of Manchester, GB*)

License  Creative Commons BY 3.0 Unported license
© Manuel López-Ibáñez

When facing realistic scenarios of automatic algorithm configuration and design, there are some questions for which our current answers appear lacking to practitioners. One practical question is how to create a set of instances representative of a problem or, alternatively, if only a small set of such instances is available, how to split the set between training and validation making sure that the training set remains representative of the target problem. Moreover, in some scenarios, we may be interested in solving a single very expensive problem instance once with the best algorithm possible. What are the strategies that automated algorithm configuration can offer in such scenarios? Another frequent question from practitioners is how to evaluate our confidence in the best configurations found, how important are the settings chosen and how sensitive are these particular settings. Some work has been done in this regard, but there are still many open questions. Finally, a more recent question is how to best integrate human interaction in the automated configuration procedure, when the target algorithm relies on a human to guide them to the optimal solution, such as in multi-criteria optimization methods that elicit preferences from decision-makers.

3.21 Building and exploiting a non-parametric problem space

Andres Munoz Acosta (*Monash University – Clayton, AU*)

License  Creative Commons BY 3.0 Unported license
© Andres Munoz Acosta

While automated algorithm selection methods have been very successful in practice, in some cases they depend on features that can be qualified as heuristics. Therefore, there are no guarantees that the representation of the problem space is efficient; or that the insights gained from the features can be turned into useful algorithms. In other words, how to exploit the features to construct useful algorithms? Perhaps the first step is to produce an efficient map of the problem space, such that the map is one-to-one. Then, we may be able to identify a transformation that would convert a new problem into a previously observed one for which we know a good –or the best– algorithm. While similar ideas exist in the literature, e.g., merging branches and nodes to make a coarse grained version of the problem, such transformations are somewhat generic. On a side note, can we find such transformations in a principled and efficient way, perhaps on-line?

References

- 1 M.A. Munoz and K.A. Smith-Miles, “Performance analysis of continuous black-box optimization algorithms via footprints in instance space”, *Evol. Comput.*, http://dx.doi.org/10.1162/EVCO_a_00194.

3.22 Automated Selection of Tree Decompositions

Nysret Musliu (TU Wien, AT)

License © Creative Commons BY 3.0 Unported license
© Nysret Musliu

Joint work of Michael Abseher, Frederico Dusberger, Nysret Musliu, Stefan Woltran
Main reference M. Abseher, F. Dusberger, N. Musliu, S. Woltran, “Improving the Efficiency of Dynamic Programming on Tree Decompositions via Machine Learning”, in Proc. of the 24th Int’l Joint Conf. on Artificial Intelligence (IJCAI 2015), pp. 275–282, AAAI Press/IJCAI, 2015.
URL <http://ijcai.org/Abstract/15/045>

Dynamic Programming (DP) over tree decompositions is a well-established method to solve problems that are in general NP-hard – efficiently for instances of small treewidth. Experience shows that DP algorithms exhibit a high variance in runtime when using different tree decompositions (TD). In fact, given an instance of the problem at hand, even decompositions of the same width might yield extremely diverging runtimes.

We propose a general method that is based on selection of the best decomposition from an available pool of heuristically generated ones. Novel features for tree decomposition are proposed and machine learning techniques are applied to select the most promising decomposition. Extensive experiments in different problem domains show a significant speedup when choosing the tree decomposition according to this concept over simply using an arbitrary one of the same width.

3.23 Feature-Based Diversity Optimization for Problem Instance Classification

Samadhi Nethmini Nallaperuma (University of Sheffield, GB)

License © Creative Commons BY 3.0 Unported license
© Samadhi Nethmini Nallaperuma

Joint work of Wanru Gao, Samadhi Nethmini Nallaperuma, Frank Neumann

Understanding the behaviour of heuristic search methods is a challenge. This even holds for simple local search methods such as 2OPT for the Traveling Salesperson problem. In this paper, we present a general framework that is able to construct a diverse set of instances that are hard or easy for a given search heuristic. Such a diverse set is obtained by using an evolutionary algorithm for constructing hard or easy instances that are diverse with respect to different features of the underlying problem. Examining the constructed instance sets, we show that many combinations of two or three features give a good classification of the TSP instances in terms of whether they are hard to be solved by 2OPT.

This research has been supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 618091 (SAGE) and by the Australian Research Council under grant agreement DP140103400.

3.24 Algorithm Selection in Music Data Analysis

Günter Rudolph (TU Dortmund, DE)

License © Creative Commons BY 3.0 Unported license
© Günter Rudolph

Joint work of Günter Rudolph, Igor Vatolkin

The analysis of signal data arising from music recordings offers many ways to apply machine learning methods. In case of classification tasks many different algorithms may be deployed which must be configured appropriately. The manual selection and parameterization of algorithmic alternatives is often necessary in most systems that realize the classification chain of musical data. The automation of this task offers the perspective of huge improvements in performance. The next years will show how much can be gained by deep neural networks that are currently built into existing systems.

3.25 Cognitive Assistant for Data Scientist

Horst Samulowitz (IBM TJ Watson Research Center – Yorktown Heights, US)

Joint work of Gregory Bramble, Maria Butrico, Andre Cunha, Elias Khalil, Udayan Khurana, Peter Kirchner, Tim Klinger, Fatemeh Nargesian, Srinivasan Parthasarathy, Chandra Reddy, Anton Riabov, Horst Samulowitz, Gerry Tesauro, Deepak Turaga

License © Creative Commons BY 3.0 Unported license
© Horst Samulowitz

A Data Scientist typically performs a number of tedious and time-consuming steps to derive insight from a raw data set. The process usually starts with data ingestion, cleaning, transformation (e.g. outlier removal, missing value imputation), then model building, and finally a presentation of predictions that align with the end-users objectives and preferences. It is a long, complex, and sometimes artful process requiring substantial time and effort especially because of the combinatorial explosion in choices of algorithms (and platforms), their parameters, and their compositions. Tools that can help automate steps in this process have the potential to accelerate the time-to-delivery of useful results, expand the reach of data science to non-experts, and offer a more systematic exploration of the available options.

This system aims at showing how automatic composition and configuration of data analytics (spanning multiple analytic platforms and packages such as R, Weka, SPSS, Apache SPARK, System ML) can offer increased insight into the data and how model selection algorithms can suggest models that are well suited to the predictive task, while respecting user preferences. Given a data set and analysis task (e.g., classification or regression) the system aims to quickly determine an appropriate combination of preprocessing steps (e.g., feature reduction or mapping) and models and platforms to achieve the users goals.

During this process the user is presented with intermediate results and insights into the data and the reasoning process itself. For example, which features are important? How well do entire classes of analytic tools perform on the data set? Are there potentially significant outliers? The user can directly interact with the process providing resource constraints (e.g., time available for training/testing) or their preference (e.g. for interpretable models). In addition, the system aims to provide basic suggestions of potentially related work that may enable the data scientist to improve results even further.

The system is constructed on top of an automated analytics composer and analytic repository which supports massively parallel execution and cross-platform execution of a wide range of high-performance analytic tools. To automatically select, compose and configure

these analytics we develop meta-learning algorithms that attempt to rapidly estimate upside performance using only subset of the available data. Furthermore, it tries to exploit structured as well as unstructured data to provide further suggestions.

3.26 From off-line to on-line feature-based parameter tuning

Marc Schoenauer (INRIA Saclay – Orsay, FR)

License © Creative Commons BY 3.0 Unported license
© Marc Schoenauer

Main reference N. Belkhir, J. Dréo, P. Savéant, M. Schoenauer, “Feature Based Algorithm Configuration: A Case Study with Differential Evolution”, in Proc. of the 14th Int’l Conf. on Parallel Problem Solving from Nature – PPSN XIV, LNCS, Vol. 9921, pp. 156–165, Springer, 2016.

URL http://dx.doi.org/10.1007/978-3-319-45823-6_15

Off-line parameter tuning based on feature computation can be achieved via the learning of an Empirical Performance Model trained on a large dataset of features x parameters, performance instances (involving huge computational cost, but this is not the point here). Given an unknown instance with computed features, optimal parameters according to the EPM can be derived. The features are of course problem-dependent, but quite often involve the computation of the objective values on a (as small as possible) set of sample points uniformly drawn from the design space – and hence can be thought of as global features. However, during the search itself, more points of the design space are sampled, and if the features are re-computed using this biased sample (or adding it to the original sample), some more “local” values of the features are obtained, that might lead to new optimal parameters according to the EPM. Very preliminary results have been obtained for continuous optimization using DE (see PPSN 2016 paper by Belkir et al.).

3.27 Cognitive Computing

Meinolf Sellmann (IBM TJ Watson Research Center – Yorktown Heights, US)

License © Creative Commons BY 3.0 Unported license
© Meinolf Sellmann

The role of IT is fundamentally shifting as our ability to collect and harness ever growing amounts of data improves. Historically an enabler of business, IT is now moving closer and closer to the heart of modern economies by informing and influencing key strategic business decisions. Human data science labor cannot keep up with the ever growing demand for decision support analytic models. The programmable era is thus coming to an end as the cognitive era of machines that practice self-orientation begins. In this presentation I invite the participants to discuss the role of meta-algorithmics in the cognitive era.

3.28 Automated generation of high-performance heuristics from flexible algorithm frameworks: Challenges and Perspectives

Thomas Stützle (Free University of Brussels, BE)

License  Creative Commons BY 3.0 Unported license
© Thomas Stützle

The design of algorithms for computationally hard problems is time-consuming and difficult for a number of reasons such as the complexity of such problems, the large number of degrees of freedom in algorithm design and the setting of numerical parameters, and the difficulties of algorithm analysis due to heuristic biases and stochasticity. In recent years, automatic algorithm configuration methods have been developed to effectively search large and diverse parameter spaces; these methods have been shown to be able to identify superior algorithm designs and to find performance improving parameter settings.

In this talk, I will shortly summarize the main recent results that we have obtained in the automatic design of hybrid stochastic local search algorithms as well as multiobjective optimizers. We show that even for problems that have received very high attention in the literature new state-of-the-art algorithms can be obtained automatically, that is, without manual algorithm tuning. I will use these recent advances to discuss informally possible directions for the future work in this direction and discuss possible challenges.

3.29 Bringing the human back in the loop

Joaquin Vanschoren (TU Eindhoven, NL)

License  Creative Commons BY 3.0 Unported license
© Joaquin Vanschoren

There has been great progress on fully-automated approaches for machine learning. Sometimes, however, human experience, intuition, or domain knowledge can prove valuable to constrain the space of solutions to explore. This is especially true when we consider the larger pipeline of data science, which also includes data wrangling, data cleaning, data integration, and model post-processing, among others. In this short talk, I would like to discuss and elicit ways to couple human expertise and machine learning to create a human-machine symbiosis. The human scientist would focus on the science (follow hunches, include more data,...) while letting the machine take care of drudge work (finding similar datasets, selecting algorithms/hyperparameters,...), thus enabling her to make informed, data-driven decisions. Meanwhile, the machine would learn from *all* the experiments run during these collaborations (with many people), and leverage what it learned from previous problems to help humans better in the future.

3.30 A Generic Bet-and-run Strategy for Speeding Up Traveling Salesperson and Minimum Vertex Cover

Markus Wagner (University of Adelaide, AU)

License © Creative Commons BY 3.0 Unported license
© Markus Wagner

Joint work of Tobias Friedrich, Timo Kötzing, Markus Wagner

Main reference T. Friedrich, T. Kötzing, M. Wagner, “A Generic Bet-and-run Strategy for Speeding Up Traveling Salesperson and Minimum Vertex Cover”, arXiv:1609.03993v1 [cs.AI], 2016.

URL <https://arxiv.org/abs/1609.03993v1>

A common strategy for improving optimization algorithms is to restart the algorithm when it is believed to be trapped in an inferior part of the search space. However, while specific restart strategies have been developed for specific problems (and specific algorithms), restarts are typically not regarded as a general tool to speed up an optimization algorithm. In fact, many optimization algorithms do not employ restarts at all.

Recently, bet-and-run was introduced in the context of mixed-integer programming, where first a number of short runs with randomized initial conditions is made, and then the most promising run of these is continued. In this article, we consider two classical NP-complete combinatorial optimization problems, traveling salesperson and minimum vertex cover, and study the effectiveness of different bet-and-run strategies. In particular, our restart strategies do not take any problem knowledge into account, nor are tailored to the optimization algorithm. Therefore, they can be used off-the-shelf. We observe that state-of-the-art solvers for these problems can benefit significantly from restarts on standard benchmark instances.

3.31 Reducing the size of large instance repositories

Markus Wagner (University of Adelaide, AU)

License © Creative Commons BY 3.0 Unported license
© Markus Wagner

Over the years, some repositories of test instances have grown significantly. Obviously, there are many more-or-less-biased ways to reduce a given set: based on algorithm performance (e.g. pick the ones where I beat my competition), based on instance features (e.g. largest 10%), and so on. Do we want to represent the distribution of the instances only, or also the density? Long story short: what is the least-biased way to reduce a given repository, and to which extent can we define the faithful (?) subset selection problem. Bonus problem: how to deal with holes in the spaces? Obviously, there is some existing work on very concrete aspects out there... how far up in generality can we go? What would generic algorithmic approaches be?

Among other, this talk gave rise to concepts like instance portfolios, marginal contribution of instances to a portfolio, and to a cycle of algorithm configuration (on instances) and instance generation (for algorithms).

3.32 Accelerating Algorithm Development

Simon Wessing (TU Dortmund, DE)

License © Creative Commons BY 3.0 Unported license
© Simon Wessing

Main reference S. Wessing, “Towards a Systematic Development Process of Optimization Methods”, arXiv:1603.00001v2 [math.OC], 2016.

URL <https://arxiv.org/abs/1603.00001v2>

Many pitfalls are lurking in algorithm engineering, and it seems that more often than not, they are not related to the solution of the mathematical problem, but to formulating the problem, implementing the algorithm, experimenting with it, and applying it to the real world. These issues typically do not get the attention of scientific research, but may severely bias its outcomes. I give examples where this has happened and try to give recommendations on how to avoid such problems in the development process, with a main focus on the planning of experiments with pre-design experiment guide sheets.

4 Breakout Sessions and Working Groups

4.1 All you ever wanted to know/ask a theoretician and All you ever wanted to know/ask a practitioner

Anne Auger (INRIA Saclay – Orsay, FR) and Carola Doerr (CNRS and University Pierre & Marie Curie – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Anne Auger and Carola Doerr

The objective of this breakout session was to gather theoreticians and practitioners to (1) better appraise the relevant questions that can or should be addressed by theoreticians for helping practitioners getting insights into the working principles of algorithm selection and configuration and (2) for practitioners to learn where already existing theoretical results could be beneficial in their research activities. The practitioners who were present are mainly working on algorithm configuration and selection while the theoreticians had mostly experience on evolutionary computation methods (including online adaptive methods). Given this difference of background, a substantial amount of time was spent on understanding the problematic on algorithm selection. More precisely, we have mostly discussed how empirical work could influence theory in algorithm configuration and selection. It has been suggested that, for example, in algorithm configuration quite often features that are seen to have a substantial impact on the performance of an algorithm are not very well understood. It could be beneficial for both theoreticians and practitioners to shed light on these effects. To this end, we have discussed potential problems. It is commonly agreed on that analyzing feature-based performance is probably out of reach for NP-hard problems like MAX-SAT and others. On the other hand, even results for much easier problems and algorithms are currently missing.

It is mentioned that not very often statistical models are used as a starting point for theoretical investigations. We discuss why this is the case (typically regarded problems are very difficult to analyze rigorously while the more easily analyzed problems are less interesting for researchers in the more empirical domains) and that theory for simple models could already be quite insightful.

During the discussion Tim Roughgarden’s work on “A PAC approach to application-specific algorithm selection” is mentioned as a theory-based study of algorithm selection problems. Furthermore, it is mentioned that in the SAT community the exchange of ideas between empirically and theoretically oriented researchers works quite well. A possible reason for this is the fact that the people agree on the problems that are analyzed and that they agree on common terminology. This is unfortunately not the case in algorithm selection and configuration where even the terminology needs to be agreed upon.

An idea emerging from this breakout session is to try algorithm selection and configuration with components that are “hand-picked” from theoreticians, e.g., by selecting only those which have been analyzed with mathematical rigor and to compare the results that one can achieve with such an approach with those being obtained without any restriction. The gap between such figures could serve as an interesting starting point for further discussions and investigations.

Acknowledgement. Besides the session chairs, the session was attended by Thomas Bäck, Benjamin Doerr, Marcus Gallagher, Wanru Gao, Holger Hoos, Frank Hutter, Lars Kotthoff, Kevin Leyton-Brown, Andres Munoz Acosta, Nysret Muliu, Frank Neumann, Marc Schoenauer and Hao Wang; we are very thankful for their valuable contributions to the discussion.

4.2 Controlled Problem Instance Generation

Marcus Gallagher (The University of Queensland – Brisbane, AU)

License  Creative Commons BY 3.0 Unported license
© Marcus Gallagher

In this flex-time session, we discussed current work, open problems and future directions in controlled (aka targeted or strategic) problem instance generation, as a key component in the evaluation of algorithm selection and configuration techniques. A photo of the whiteboard after the session is attached.

Conceptually, for combinatorial, discrete, continuous (or mixed?) black-box optimization problems, we are interested in the performance of algorithm instances (e.g. from an algorithm selection or configuration technique) over some set of problems. The set of all problems may not be interesting (as implied by the No Free Lunch Theorems), since performance (for many definitions) is equal, on average. However many of these problems are uninteresting (e.g. “white noise”): we are really interested in the subspace of problems with some sort of exploitable structure, and/or those with relevance to real-world problems. It is currently unclear to what extent commonly used benchmark problem sets (e.g. BBOB for continuous problems) represent the set of “interesting” problems.

For algorithm evaluation, it is desirable to have good coverage of the set of “interesting” problems when performing experiments. If different algorithms perform well/poorly on different types of problems, good coverage is important to get a clear picture of this. However this raises the question of what is meant by “types of problems”. A good part of our discussion was consequently about problem features – since we need ways of measuring the characteristics of problems to identify or measure “problem type”. Good features should help us gain a better understand of algorithm performance. It might also be possible to perform better experiments if test problems can be generated which vary smoothly with respect to problem features.

Several different possibilities were discussed regarding controlled problem instance generation:

- Evolving problems using Genetic Programming. Here we have a symbolic representation for the problems – representation is clearly a general, important issue.
- Use of a heuristic search algorithm (e.g. 1+1-EA) to generate problems with some desirable property in feature space.
- Finding real-world representative problems.
- Using surrogate or generative models.
- Blending known functions.

Issues around “feature selection” were also discussed. How many features are needed and how to select them is a issue from Machine Learning. When features are based on sampling the fitness landscape, the sampling technique and size becomes important. The curse of dimensionality suggests that more features require a much larger sample size to support effective estimation.

The session identified many interesting and important issues that would make fruitful research directions.

Acknowledgement. Besides the session chairs, the session was attended by Wanru Gao, Pascal Kerschke, Lars Kotthoff, Andres Munoz Acosta, Samadhi Nethmini Nallaperuma, Mike Preuß, and Heike Trautmann; we are very thankful for their valuable contributions to the discussion.

4.3 Describing / Characterizing Landscapes of multiobjective Optimization Problems

Pascal Kerschke (Universität Münster, DE)

License  Creative Commons BY 3.0 Unported license
© Pascal Kerschke

This is a summary of the breakout session on *Describing / Characterizing Landscapes of multiobjective Optimization Problems*, which was held at the *Dagstuhl Seminar 16412 on Automated Algorithm Selection and Configuration* on October 13, 2016.

Initiated by recent research results, which started to characterize multiobjective optimization problems, this breakout session was originally intended to discuss the following four issues:

1. What are characteristics / landmarks / properties of a multiobjective landscape?
2. How can we measure the interaction of the objectives (in addition to simply using indicators)?
3. What could be (cheaply computable) features of a multiobjective landscape?
4. Are there differences across the different domains (continuous, discrete, TSP, etc.)?

However, during the roughly 60-70 minutes of brainstorming, the approx. 10 participants of this session actually re-defined this session. So, we started with some points of the list from above and then the discussion took off..

Measuring the Similarity Between the Objectives

In a first topic, the participants discussed ways to measure the interaction of the objectives (i.e., issue 2 from the original questions). The initial idea was to measure the covariances and/or correlations between the objectives and to somehow estimate the underlying multivariate distribution. In the single-objective scenario, there seem to exist related approaches, called *density of states*. But what would such a similarity information tell us?

Another idea was to have a look at the contour lines of the points in the objective space and somehow use that information to differ between local and global optima.

Detection of Disconnected Pareto Fronts

Based on the idea of analyzing the Pareto fronts, we came to a discussion on whether it is possible to detect disconnected Pareto fronts. One possible solution was to have a look at the objective-wise gradients and see whether they point in opposite directions. Another idea was to apply a clustering approach to the points of the objective space and use the number of found (non-dominated) clusters as representatives of disconnected (global) fronts.

Aggregating the Objectives

In a next approach, Carlos Fonseca introduced us to the basic idea of his PhD-thesis, in which he represented the multiobjective landscape by a single-objective one.

Inspired by the idea of reducing the multiobjective problem to a single-objective one, we came up with the idea of computing the landscape features objective-wise and aggregating them afterwards. But then the question would be how to aggregate that information in a meaningful way and what does it tell us?

Also, is it really useful to spend the same amount of function evaluations for each of the objectives? Maybe some of the objectives are rather easy and thus cheap to compute, whereas others are more complex. Therefore, one could also compute surrogate models for each of the objectives and then use these models to compute numerous (hopefully representative) landscape features of the original landscape.

Approaching the Problem From Different Angles

For some reason, we mainly focussed on the objective space and tried to think of approaches on how to characterize the information that's hidden in there. One idea was to make use of features from the TSP domain. So for instance, one could have a look at the points (= cities) in the objective space and compute features based on their distance matrix, MST or convex hull.

There was also the idea to analyze the path of the (TSP) features across the iterations of the optimization algorithms. That is, we run the optimization algorithm and for each generation, we use the current population to compute the (TSP) features and then analyze how they change over the course of time.

Inspired by the cat-like shapes of the objective space (sketched on the whiteboard), there were also two more ideas coming up:

1. There should be new benchmark functions; problems with specific shapes such as the cat-like shape that was sketched on the board.
2. We could (at least for now) skip the landscape features and see how other (promising) approaches perform. So, one idea was to consider the shape of the sampled objective space as image and then use a deep-learning neural network for performing algorithm

selection on the multiobjective problems. This might not give us direct insights into understanding the characteristics of the problems themselves, but they would at least provide a solid base line for feature-based algorithm selection models. And we could afterwards use these results to find problems for which the algorithms behaved differently and then use that information to develop features which could be more promising for describing these differences.

Acknowledgement. Besides the session chairs, the session was attended by Michael Emmerich, Carlos M. Fonseca, Marcus Gallagher, Carlos Ignacio Hernández Castellanos, Frank Hutter, Manuel López-Ibáñez, Andres Munoz Acosta, Heike Trautmann, Markus Wagner, Hao Wang and Simon Wessing; we are very thankful for their valuable contributions to the discussion.

4.4 Portfolio-based Methods for Algorithm Selection

Kevin Leyton-Brown (University of British Columbia – Vancouver, CA)

License  Creative Commons BY 3.0 Unported license
© Kevin Leyton-Brown

We had a breakout session on portfolio based methods for algorithm selection. The topics we considered were divided into best practices and hurdles. In the former category, here are questions the group considered:

- Which methods do you prefer and why?
- Have you been involved in successful applications
- What do you consider the role of features?
- Describe your experiences with aslib
- Best approaches for deciding what to put in the portfolio
- Good ideas you think others don't know about
- How to detect incorrect algorithm behavior
- What statistical methods are necessary to ensure validity of results?
- How do things change when selecting parallel solvers?

Here are the questions we considered regarding hurdles:

- What doesn't work?
- Do we still need selection in an increasingly parallel world?
- What interesting findings have you not published?
- How important is fancy machine learning? (And, if it's important, what fancy methods do you like?)
- How is selection connected to PIAC (they are the same; selection is subsumed; ...?)

Finally, we discussed the big questions the field should grapple with next. Here is a list of topics identified by the group.

- Parallelism. What if you run everything in parallel?
- Tradeoff between restarts and parallel runs
- How to find many, truly complementary algorithms?
- Is there a sense in which two algorithms can be said to be complementary "always"?
- What can we learn from ML literature on ensemble methods?
- Methods for regularizing portfolios

- Tradeoff in black-box continuous optimization between gathering data for use in features and using the same data in the optimization itself
- PIAC: predicting parameter values from continuous space

4.5 What can we learn from algorithm selection data?

Marius Lindauer (Universität Freiburg, DE) and Lars Kotthoff (University of British Columbia – Vancouver, CA)

License © Creative Commons BY 3.0 Unported license
© Marius Lindauer and Lars Kotthoff

Main reference B. Bischl, P. Kerschke, L. Kotthoff, M. Lindauer, Y. Malitsky, A. Frechétte, H. Hoos, F. Hutter, K. Leyton-Brown, K. Tierney, J. Vanschoren, “ASlib: A Benchmark Library for Algorithm Selection”, *Artificial Intelligence*, Vol. 237, pp. 41–58, Elsevier, 2016.

URL <http://dx.doi.org/10.1016/j.artint.2016.04.003>

Selecting a well-performing algorithm for a given problem instance (e.g., a combinatorial problem or machine learning data set) often substantially improves the performance compared to always using the same algorithm. The automation of this process is called automatic algorithm selection [4] which is often implemented by using machine learning models requiring a lot of training data to get decent predictions. This training data mainly consist of two required matrices, i.e., the performance of each algorithm on each instance, and the instance features for each instance. To reduce the burden on algorithm selection developers to collect these data and to provide standardized data for comparison of algorithm selectors[3], the algorithm selection library ASlib [1] was created.

Besides doing algorithm selection experiments on the data in ASlib, the question of the breakout session was what can we further do with these data to get more insights into the algorithm selection problem to solve, and to get insights why and on which instances an algorithm selector performs well.

A first step in this direction is already done in the exploratory data analysis (EDA) provided on the ASlib website². A user gets insights in the performance distributions of each algorithm, performance correlation of pairs of algorithms and distributions of instance features. To extend this data-driven overview, we discussed further plots to show portfolio contributions of algorithms [2], statistical significance tests, feature importance and cost of computing instance features.

Up to now, papers on algorithm selectors often only report how well they perform on all instances but we lack some detailed analyses which instances they perform well on and where they fail to select a well-performing algorithm. To this end, we discussed that it would be nice to have interactive scatter plots showing the performance of the single best algorithm and an algorithm selector on each instance. Furthermore, a new idea is to train an easy-to-interpret decision tree to classify on which instances the algorithm selector performs well. Such plots could be in principle also automatically generated, if ASlib would allow to upload results from algorithm selection experiments (similar to OpenML [5]).

An further open question is how to automatically pass information extracted by the EDA to an algorithm selector. For example, if the EDA already figured out that some algorithms are dominated and not needed for an algorithm portfolio, or which instance features are

² <http://www.aslib.net>

important, the information could be directly exploited for training an algorithm selector. Right now, this process is still mainly manual.

Currently, ASlib provides 17 algorithm selection benchmarks, all from hard combinatorial problems. Whether these benchmarks can be called real-world benchmarks is unknown. However, a mid-term goal of ASlib is include even more diverse benchmarks, for example from the continuous optimization community and from the meta-learning community in machine learning.

In summary, ASlib was well-received and provides a lot of untouched potential to learn more about algorithm selection data and the behavior of algorithm selectors on them.

Acknowledgement. Besides the session chairs, the session was attended by Wanru Gao, Holger Hoos, Nysret Musliu, Samadhi Nethmini Nallaperuma, Marc Schoenauer and Markus Wagner; we are very thankful for their valuable contributions to the discussion.

References

- 1 B. Bischl, P. Kerschke, L. Kotthoff, M. Lindauer, Y. Malitsky, A. Frech ette, H. Hoos, F. Hutter, K. Leyton-Brown, K. Tierney, and J. Vanschoren. ASlib: A benchmark library for algorithm selection. *Artificial Intelligence*, pages 41–58, 2016.
- 2 A. Fr echette, L. Kotthoff, T. Michalak, T. Rahwan, H. Hoos, and K. Leyton-Brown. Using the shapley value to analyze algorithm portfolios. In D. Schuurmans and M. Wellman, editors, *Proceedings of the Thirtieth AAAI conference ON Artificial Intelligence*, pages 3397–3403. AAAI Press, 2016.
- 3 M. Lindauer, H. Hoos, F. Hutter, and T. Schaub. Autofolio: An automatically configured algorithm selector. *Journal of Artificial Intelligence Research*, 53:745–778, August 2015.
- 4 J. Rice. The algorithm selection problem. *Advances in Computers*, 15:65–118, 1976.
- 5 J. Vanschoren, J. van Rijn, and B. Bischl. Taking machine learning research online with openml. In *Proceedings of the 4th International Workshop on Big Data, Streams and Heterogeneous Source Mining (BigMine)*, volume 41 of *JMLR Workshop and Conference Proceedings*, pages 1–4. JMLR.org, 2015.

4.6 (ELA features for) multimodal optimization

Mike Preu  (Universit t M nster, DE)

License   Creative Commons BY 3.0 Unported license
  Mike Preu 

Main reference K. Kawaguchi, “Deep Learning without Poor Local Minima”, arXiv:1605.07110v3 [stat.ML], 2016.
URL <https://arxiv.org/abs/1605.07110v3>

This session was intended for brainstorming on possible new Exploratory Landscape Analysis (ELA) features that are especially well suited for extracting knowledge from multimodal optimization problems. However, the discussion centered much more on what the basic properties of multimodal problems are, or even more general, how multimodal optimization itself is defined. As the term is relatively young, there are several opinions on the basic ideas and the need for better definitions was expressed. One example is the optimum definition for ridge functions: is it a set of points, or an areal structure? This even holds true for plateaus, which may be considered as a large set of optima or a single optimum (however, this would be far from a mathematically rigid definition). Detection of a plateau seems to be easy for a human, and more difficult (but probably possible) automatically, even in high dimensions. It is also unclear how common such optimum types are for real world problems, but it is expected that they are of some importance.

Generally, we agreed that we need to rigidly define multimodality, funnel, the whole vocabulary used for this kind of optimization, and that based on that, we need some generally accepted measures. For example, how can we express the multimodality of a problem as compared to another in a value? And how the robustness of peaks?

Another general question that was discussed to some extent was the one for the type of peaks we actually want to detect. Are we satisfied with good local optima? How many? The currently employed benchmarks concentrate on multi-global optimization, but the group considers this rather a special case than of general interest. Additionally, taking into account the recent developments in deep learning, it may be even satisfactory to find good saddle points (good in this context means some that are optima in most dimensions but saddle points in few)? Can we derive benchmark functions with saddle points? The positivity of the Hessian would be a good indicator. We expect interesting results from optimization experiments on such problems as saddle points are difficult to escape. At least, an evolutionary algorithm would probably be slowed down and it could happen that this prematurely triggers termination criteria, making the problem even the more difficult.

The saddle points discussion was based on this work: Deep Learning without Poor Local Minima by Kenji Kawaguchi: <https://arxiv.org/abs/1605.07110>

Acknowledgement. Besides the session chairs, the session was attended by Carlos M. Fonseca, Marcus Gallagher, Pascal Kerschke, Andres Munoz Acosta, Heike Trautmann and Simon Wessing; we are very thankful for their valuable contributions to the discussion.

4.7 Online and Adaptive Methods

Marc Schoenauer (INRIA Saclay – Orsay, FR)

License  Creative Commons BY 3.0 Unported license
© Marc Schoenauer

Preliminary

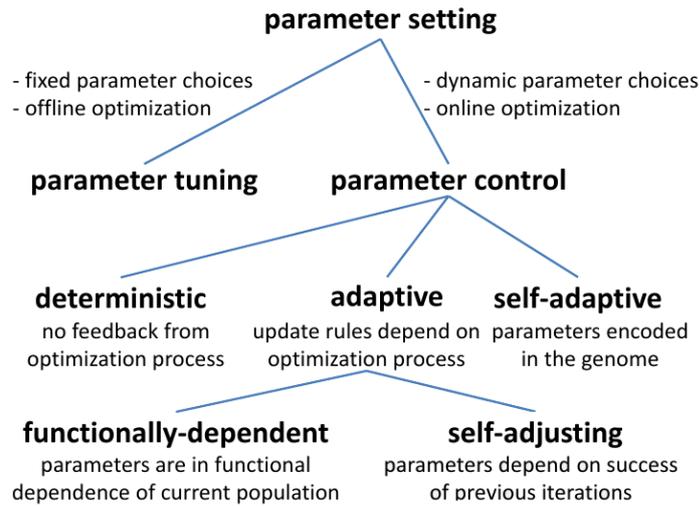
This document is the main outcome of the breakout session about *Online and Adaptive Methods* that took place during Dagstuhl Seminar *Algorithm Selection and Configuration* – October 10-15.

4.7.1 Monday Flex group meeting

The whole discussion started when Thomas Stützle heavily criticized the long-known diagram, originally proposed by Eiben et al. in their 1999 paper in IEEE TEC [1]. Figure 1 is an adaptation of the initial diagram, taken from [2].

Some mathematical framework was proposed to the discussion by Anne Auger, quickly supported by Carlos Fonseca . . . as this was very close to his own proposal, that he presented on the following Tuesday morning during his short talk about dynamical systems [3]. The discussion then tried to instantiate known instances of parameter setting withing this model, modifying the model itself when necessary – and partly continued during the following days.

At the end of the Dagstuhl seminar, all participants had agreed on the following proposal, that is now offered to the whole community in the hope it can be adopted and improved in order to somehow represent all know instances of parameter setting.



■ **Figure 1** Doerr & Doerr’s version [2] of Eiben et al.’s classification diagram [1].

4.7.2 The proposal

- x_t is the population at large (points of the search space). Thus here, $x_t \in \Pi$, space of populations of search points. Note that this includes algorithms with archives, but not algorithms that make use of all population from the very beginning at all steps, as pointed out by Michael).
- σ_t is a set of parameters that are used to update the population, and are themselves updated (or not).
- u_t (and v_t) are uniformly and independently generated random numbers

4.7.2.1 Static parameter setting

$$\begin{cases} x_{t+1} = F(x_t, \sigma_t, u_t) \\ \sigma_{t+1} = \sigma_0 \end{cases} \quad (1)$$

Some people (lead by Anne Auger and Benjamin Dörr) suggested to replace u_t with u_{t+1} in this equation. Opponents (Carlos Fonseca, Marc Schoenauer ...) agreed in the end that it is a minor and formal detail.

4.7.2.2 Non-adaptive / feedback-free parameter setting

$$\begin{cases} x_{t+1} = F(x_t, \sigma_t, u_t) \\ \sigma_{t+1} = G(\sigma_t, t, v_t) \end{cases} \quad (2)$$

4.7.2.3 Adaptive parameter setting

$$\begin{cases} x_{t+1} = F(x_t, \sigma_t, u_t) \\ \sigma_{t+1} = G(x_t, \sigma_t, u_t) \end{cases} \quad (3)$$

with 2 subsets:

4.7.2.4 Functionally dependent

$$\begin{cases} x_{t+1} &= F(x_t, \sigma_t, u_t) \\ \sigma_{t+1} &= G(F(x_t, \sigma_t, u_t)) \end{cases} \quad (4)$$

Note: the second equation could be written as $\sigma_{t+1} = G(x_{t+1})$ but is kept that way to be consistent with the usual form of dynamical system definition.

4.7.2.5 Self-adjusting

This is the part (name and equations) that was most heavily discussed – and no real consensus was reached.

$$\begin{cases} x_{t+1} &= F(x_t, \sigma_t, u_t) \\ \sigma_{t+1} &= G(x_t, u_t) \end{cases} \quad (5)$$

4.7.3 Self-adaptive

The formal definitions above lead to unforeseen difficulties when it came to address self-adaptive properties. Two proposals were made:

- *Hide* the mutation parameters in the definition of the state space, championed by Marc Schoenauer;
- Create some intermediate step to reflect self- property, proposed by Carlos Fonseca.

4.7.3.1 Changing the State Space

The dynamical system formulation is cool – it allows mathematical proofs depending on properties of F and G (see previous work from Anne Auger, and [3] in this document). It should be preserved/extended by relaxing the definition of the first part of the state space where x_t 'lives' (thus meeting the more general formulation that Anne had initially proposed :-)

If you allow the x_t in the above dynamical systems definitions to be something else than a population of search points, this could cover both issues of EDAs and self-adaptive:

- For EDAs, it should be the distribution d_t that evolves. Then function F describes how the distribution is updated (including sampling), and σ the parameters of this sampling/adaptation.
- For self-adaptive algorithms, it should be $\mathcal{S} \otimes \times$, i.e. representing points of the search space to which are attached some parameters (e.g., the mutation parameters in self-adaptive ES, the crossover bit for Spear's GA, etc).

4.7.3.2 Adding Intermediate Stages

For self-adaptation:

$$\begin{cases} \sigma_{t+\frac{1}{2}} &= f_1(\sigma_t, u_t) \\ x_{t+\frac{1}{2}} &= f_2(x_t, \sigma_{t+\frac{1}{2}}, u_t) \\ x_{t+1} &= f_3(x_{t+\frac{1}{2}}) \\ \sigma_{t+1} &= f_4(x_{t+\frac{1}{2}}, \sigma_{t+\frac{1}{2}}) \end{cases} \quad (6)$$

and for self-adjusting:

$$\begin{cases} x_{t+\frac{1}{2}} = f_1(x_t, \sigma_t, u_t) \\ x_{t+1} = f_2(x_{t+\frac{1}{2}}) \\ \sigma_{t+\frac{1}{2}} = f_3(x_{t+\frac{1}{2}}, \sigma_t) \\ \sigma_{t+1} = f_4(x_t, x_{t+\frac{1}{2}}, \sigma_{t+\frac{1}{2}}) \end{cases} \quad (7)$$

4.7.3.3 Other issues with self-adaptive setting

- Does self-adaptation really work? The nomenclature is unclear, there are differences in opinions about what's adaptive and what's not (see Section 4.7.2).
- The performances of self-adaptation need to be compared to properly-tuned algorithms. In particular
 - Ultimate test: Can self-adaptive method recover the theoretically-best schedule?
 - Two mandatory baselines for all comparisons: random choice, and oracle.

References

- 1 A.E. Eiben, R. Hinterding, and Z. Michalewicz. Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* 3(2): 124–141, 1999.
- 2 B. Doerr and C. Doerr. Optimal Parameter Choices Through Self-Adjustment: Applying the 1/5-th Rule in Discrete Settings. In S. Silva and A.I. Esparcia-Alcázar, editors, *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2015*, pages 1335–1342, 2015.
- 3 C. Fonseca. *Optimisation Algorithm Design: A Control Engineering Perspective*. Dagstuhl Seminar on Automated Algorithm Selection and Configuration, 2016.

4.8 Real-world Applications of Meta-Algorithmics

Meinolf Sellmann (IBM TJ Watson Research Center – Yorktown Heights, US)

License  Creative Commons BY 3.0 Unported license
© Meinolf Sellmann

We discussed industrial and non-profit applications of meta-algorithmics that can be roughly grouped in three different categories:

4.8.1 Automated Model Lifecycle Management

The First and biggest application area we see is the automatic creation, adaptation, and risk management of data science model. This increasingly becomes a need as the availability of more data allows for individualization of predictive and prescriptive models which are deployed and dynamically changing environments. Real-world examples for this technology are, e.g.

1. In Education, the individualization of learning, motivation, content selection, and exercise selection.
2. In Heavy Industries, the management of physical assets.
3. In Medicine, the personalization of healthcare.
4. In Transportation, the provisioning of detailed forecasting models of transportation times in different regions, cities and for different modes of transportation.

In all these domains, the assurance of data science technology will be key to promote wide-spread adoption.

4.8.2 One-of strategic decision support

The second area of application are data science problems that are singular and strategic rather than operational and tactical. Meta-algorithmics can help here to bridge the gap between predictive and prescriptive modeling by automatically providing the ability to phenomenologically study uncertainties in the prediction of predictive models that provide the input for down-stream prescriptive analytics. Moreover, off-line preparation can help learn adaptive control strategies that help guide the decision process as a strategic decision making event unfolds. One example where such a technology could be deployed is in the management of disasters.

4.8.3 Combinatorial Design

Finally, we discussed the problem of assembling given parts to a whole that can be expected to match up with a given set of requirements. For example, think of a team of experts that needs to be assembled to work on a particular project. Both experts can project are described with certain features, and we need to decide which team configuration will have the best outlook to handle the project well. For such a scenario, portfolio selection techniques are applicable. Moreover, there is a research demand here regarding the design and automatic generation of predictive models that predict team performance.

4.9 Pitfalls and Best Practices for Algorithm Configuration

Marius Lindauer and Frank Hutter (Universität Freiburg, DE)

License  Creative Commons BY 3.0 Unported license
© Marius Lindauer and Frank Hutter

Automatic algorithm configuration [4] helps developers and users of all types of algorithms to tune their parameters (e.g., options of search heuristics or hyperparameters of machine learning algorithms). This can often substantially improve performance (e.g., running time or prediction error). Applying and comparing automatic algorithm configuration tools (such as *ParamILS* [4], *irace* [8], *SMAC* [3] and *GGA* [1]) is related to empirical benchmarking and can include many subtle pitfalls, even if reliable benchmark libraries such as *AClib* [6] are used. In the following, we summarize the discussion on this topic in a breakout session at the Dagstuhl seminar “Automated Algorithm Selection and Configuration”.

A common mistake in tuning parameters (also in manual tuning and development of algorithms) is to optimize parameters on the same instances that are used later to evaluate their performance. This can lead to overoptimistic performance estimates and over-tuning effects [5]. To avoid this problem, we recommend to first split the available instances into a training and test set, using the training instances for tuning and the test instances to report the performance on. Using an outer cross-validation would estimate the individual performances with lower variance, but it is typically infeasible in algorithm configuration since each single algorithm configuration run is already very expensive. Also, if the computational budget allows for more than one algorithm configuration experiment, we recommend to rather

run experiments on k algorithm configuration benchmarks than a k -fold cross-validation on a single one; this helps to also capture differences across benchmarks.

Related to performance assessment in general, measuring running time can have subtle problems; e.g., it can be influenced by noise induced by other processes running on the same machine. An alternative to measuring running time could be measuring MEMS [7], i.e., the number of memory accesses. MEMS can be measured with a very fine-grained resolution and therefore allow to precisely measure the performance of fast algorithms. To report running time in a publication, an initial study would be necessary to find a mapping from MEMS to running time (which should be a simple linear model).

A wide variety of pitfalls is related to the target algorithm being optimized and the *wrapper* around it (a communication layer between the algorithm configuration procedure and the target algorithm to be optimized). For example, some target algorithms can measure their own running time, but we have experienced that some can also return negative running time. Another example is that some users/configurators blindly optimize for running time without checking that the target algorithm has properly returned; since many algorithms have some bugs, this would often lead to optimizing the running time to crash. Therefore, we recommend to use a uniform and robust wrapper which also handles the running time measurement and resource limitations (e.g., running time or memory limits); our own solution to this is a generic Python wrapper that is easy to instantiate for a given algorithm: <https://github.com/mlindauer/GenericWrapper4AC>.

Further pitfalls are related to the parameter configuration space defined by value bounds for all parameters of a target algorithm. In order to open up the greatest performance potential, our general recommendation is to include as many meaningful parameters as possible to explore within a fixed computational budget, and to also choose their ranges large enough to prevent most of human bias. However, adding very large bounds (e.g., full 32bit integer ranges) can make it very hard for configurators to find a well-performing parameter configuration. Another pitfall is to add parameters to the configuration space that change the semantics of the algorithm or performance metric; e.g., optimizing the solution quality gap of the mixed-integer programming (MIP) solver *CPLEX* will drastically reduce the running time, but the resulting CPLEX configuration may only return poor solutions of the given MIP problems.

Many other issues were touched on in the session that would go beyond the scope of this short summary. Overall, when working with algorithms, many things can go wrong. As we use automated methods, even more things tend to go wrong. Therefore, it is important to be aware of common pitfalls and best practices to avoid them. In the upcoming book on “Empirical Algorithmics”, Holger Hoos [2] lists many useful best practices related to empirical benchmarking. Katharina Eggensperger, Marius Lindauer, and Frank Hutter are currently working on an article that describes best practices and pitfalls in algorithm configuration in more detail.

Acknowledgement. Besides the session chairs, the session was attended by Aymeric Blot, Wanru Gao, Holger Hoos, Laetitia Jourdan, Lars Kotthoff, Manuel López-Ibáñez, Nysret Musliu, Günter Rudolph, Marc Schoenauer, Thomas Stützle and Joaquin Vanschoren; we are very thankful for their valuable contributions to the discussion.

References

- 1 C. Ansótegui, Y. Malitsky, M. Sellmann, and K. Tierney. Model-based genetic algorithms for algorithm configuration. In Q. Yang and M. Wooldridge, editors, *Proceedings of the*

- 25th International Joint Conference on Artificial Intelligence (IJCAI'15)*, pages 733–739, 2015.
- 2 Holger H. Hoos. *Empirical Algorithmics*. Cambridge University Press, 2017. to appear.
 - 3 F. Hutter, H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In C. Coello, editor, *Proceedings of the Fifth International Conference on Learning and Intelligent Optimization (LION'11)*, volume 6683 of *Lecture Notes in Computer Science*, pages 507–523. Springer-Verlag, 2011.
 - 4 F. Hutter, H. Hoos, K. Leyton-Brown, and T. Stützle. ParamILS: An automatic algorithm configuration framework. *Journal of Artificial Intelligence Research*, 36:267–306, 2009.
 - 5 F. Hutter, H. Hoos, and T. Stützle. Automatic algorithm configuration based on local search. In R. Holte and A. Howe, editors, *Proceedings of the Twenty-second National Conference on Artificial Intelligence (AAAI'07)*, pages 1152–1157. AAAI Press, 2007.
 - 6 F. Hutter, M. López-Ibáñez, C. Fawcett, M. Lindauer, H. Hoos, K. Leyton-Brown, and T. Stützle. Aclib: a benchmark library for algorithm configuration. In P. Pardalos and M. Resende, editors, *Proceedings of the Eighth International Conference on Learning and Intelligent Optimization (LION'14)*, *Lecture Notes in Computer Science*. Springer-Verlag, 2014.
 - 7 Donald E. Knuth. *The Art of Computer Programming, Volume IV*. Addison-Wesley, 2011.
 - 8 M. López-Ibáñez, J. Dubois-Lacoste, T. Stützle, and M. Birattari. The irace package, iterated race for automatic algorithm configuration. Technical report, IRIDIA, Université Libre de Bruxelles, Belgium, 2011.

4.10 Multiobjective Optimisation Algorithm Selection and Configuration

Carlos M. Fonseca (University of Coimbra, PT) and Manuel López-Ibáñez (Univ. of Manchester, GB)

License © Creative Commons BY 3.0 Unported license
© Carlos M. Fonseca and Manuel López-Ibáñez

This breakout session, which was held on October 11, 2016, aimed at discussing Algorithm Selection and Configuration in Multiobjective Optimisation (MO) contexts. In particular, the following topics were proposed:

1. Selection and configuration of multiobjective optimisation algorithms
2. Algorithm selection and configuration under multiple performance criteria

Automated configuration of multiobjective optimisation algorithms was considered first. It was noted that the outcome of a MO optimisation run is often a set of non-dominated solutions, which makes it difficult to compare such outcomes directly. In the literature, two approaches have been proposed to assess the performance of MO algorithms: quality indicators and the attainment function. Quality indicators map non-dominated point sets onto real values, and make it easy to tune MO algorithms with existing automated configuration tools. However, the choice of quality indicator may influence the tuning process considerably, since different quality indicators may disagree about which of two outcomes is best. It was pointed out that this is particularly noticeable as the number of objectives increases. In contrast, the attainment function approach deals with the distribution of non-dominated point sets directly, and allows the performance of two different algorithms to be compared (to an extent) by means of hypothesis tests, although their power is perceived to be low.

Additionally, it is still not clear how attainment-function based comparisons can be performed when several benchmark problems are used for tuning.

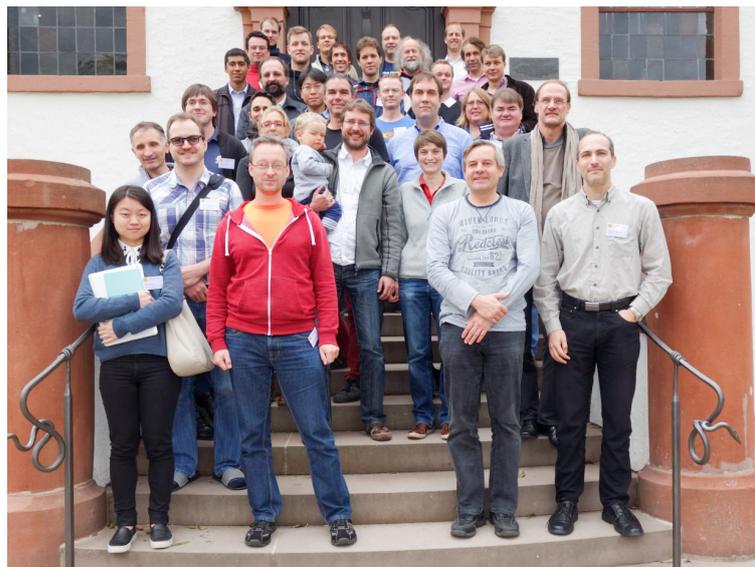
The availability of multiple quality indicators lead to the discussion of the second topic. Unless there is a clear preference for a given indicator, configuring algorithms to perform well with respect to several indicators follows naturally. Questions then arise on how to aggregate information from different indicators, especially when they are conflicting. Using quality indicators at a higher level to aggregate different quality-indicator values was suggested. Another instance of multiple configuration criteria are the runtime and solution-quality views of performance. It was argued that, by including the time at which individual solutions are found in a run as an additional objective, the resulting augmented sets of non-dominated solutions characterise the *anytime* behaviour of the corresponding algorithms. Tuning for anytime performance would then be implemented by applying quality indicators, as before. The issue of how to measure runtime (computing time versus number of function evaluations, for example) was also considered an important issue, with practical effects on configuration results.

Finally, it was felt that the number of established multiobjective benchmark problems is still very limited, despite on-going efforts to address that issue, and that there is insufficient understanding of what multiobjective problem features are relevant, and how their presence may affect algorithm performance. This discussion was continued in another breakout session on the characterisation of the landscapes of multiobjective optimisation problems. Other topics that were identified, but could not be discussed for lack of time, include the tuning of interactive multiobjective optimisation algorithms and the interplay between preference articulation and algorithm selection and configuration.

Acknowledgement. Besides the session chairs, the session was attended by Aymeric Blot, Michael Emmerich, Carlos M. Fonseca, Carlos Ignacio Hernández Castellanos, Laetitia Jourdan, Pascal Kerschke, Marius Lindauer, Manuel López-Ibáñez, Samadhi Nethmini Nallaperuma, Thomas Stützle, Heike Trautmann, Markus Wagner and Simon Wessing; we are very thankful for their valuable contributions to the discussion.

Participants

- Anne Auger
INRIA Saclay – Orsay, FR
- Thomas Bäck
Leiden University, NL
- Aymeric Blot
INRIA Lille, FR
- Benjamin Doerr
Ecole Polytechnique –
Palaiseau, FR
- Carola Doerr
CNRS and University Pierre &
Marie Curie – Paris, FR
- Michael Emmerich
Leiden University, NL
- Carlos M. Fonseca
University of Coimbra, PT
- Tobias Friedrich
Hasso-Plattner-Institut –
Potsdam, DE
- Marcus Gallagher
The University of Queensland –
Brisbane, AU
- Wanru Gao
University of Adelaide, AU
- Carlos Ignacio Hernández
Castellanos
CINVESTAV – Mexico, MX
- Holger H. Hoos
University of British Columbia –
Vancouver, CA
- Frank Hutter
Universität Freiburg, DE
- Laetitia Jourdan
INRIA Lille, FR
- Pascal Kerschke
Universität Münster, DE
- Lars Kotthoff
University of British Columbia –
Vancouver, CA
- Kevin Leyton-Brown
University of British Columbia –
Vancouver, CA
- Marius Lindauer
Universität Freiburg, DE
- Manuel López-Ibáñez
Univ. of Manchester, GB
- Andres Munoz Acosta
Monash University –
Clayton, AU
- Nysret Musliu
TU Wien, AT
- Samadhi Nethmini
Nallaperuma
University of Sheffield, GB
- Frank Neumann
University of Adelaide, AU
- Mike Preuß
Universität Münster, DE
- Günter Rudolph
TU Dortmund, DE
- Horst Samulowitz
IBM TJ Watson Research Center
– Yorktown Heights, US
- Marc Schoenauer
INRIA Saclay – Orsay, FR
- Meinolf Sellmann
IBM TJ Watson Research Center
– Yorktown Heights, US
- Thomas Stütze
Free University of Brussels, BE
- Heike Trautmann
Universität Münster, DE
- Joaquin Vanschoren
TU Eindhoven, NL
- Markus Wagner
University of Adelaide, AU
- Hao Wang
Leiden University, NL
- Simon Wessing
TU Dortmund, DE



Universality of Proofs

Edited by

Gilles Dowek¹, Catherine Dubois², Brigitte Pientka³, and Florian Rabe⁴

1 INRIA & ENS Cachan, FR, gilles.dowek@ens-cachan.fr

2 ENSIIE – Evry, FR, catherine.dubois@ensiie.fr

3 McGill University – Montreal, CA, bpientka@cs.mcgill.ca

4 Jacobs University Bremen, DE, f.rabe@jacobs-university.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16421 *Universality of Proofs* which took place October 16–21, 2016.

The seminar was motivated by the fact that it is nowadays difficult to exchange proofs from one proof assistant to another one. Thus a formal proof cannot be considered as a *universal* proof, reusable in different contexts. The seminar aims at providing a comprehensive overview of the existing techniques for interoperability and going further into the development of a common objective and framework for proof developments that support the communication, reuse and interoperability of proofs.

The seminar included participants coming from different fields of computer science such as logic, proof engineering, program verification, formal mathematics. It included overview talks, technical talks and breakout sessions. This report collects the abstracts of talks and summarizes the outcomes of the breakout sessions.

Seminar October 16–21, 2016 – <http://www.dagstuhl.de/16421>

1998 ACM Subject Classification Semantics / Formal Methods, Verification / Logic

Keywords and phrases Formal proofs, Interoperability, Logical frameworks, Logics, Proof formats, Provers, Reusability

Digital Object Identifier 10.4230/DagRep.6.10.75

1 Executive Summary

Gilles Dowek

Catherine Dubois

Brigitte Pientka

Florian Rabe

License  Creative Commons BY 3.0 Unported license
© Gilles Dowek, Catherine Dubois, Brigitte Pientka, and Florian Rabe

Proof systems are software systems that allow us to build formal proofs, either interactively or automatically, and to check the correctness of such proofs. Building such a formal proof is always a difficult task – for instance the Feit-Thompson odd order theorem, the CompCert verified C compiler, the seL4 verified operating system micro-kernel, and the proof of the Kepler conjecture required several years with a medium to large team of developers to be completed. Moreover, the fact that each of these proofs is formalized in a specific logic and the language of a specific proof tool is a severe limitation to its dissemination within the community of mathematicians and computer scientists. Compared to many other branches of



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

Universality of Proofs, *Dagstuhl Reports*, Vol. 6, Issue 10, pp. 75–98

Editors: Gilles Dowek, Catherine Dubois, Brigitte Pientka, and Florian Rabe



Dagstuhl Reports

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

computer science, for instance software engineering, we are still very far from having off-the-shelf and ready-to-use components, “proving in the large” techniques, and interoperability of theory and systems. However, several teams around the world are working on this issue and partial solutions have been proposed including point-to-point translations, proof standards, and logical frameworks. Yet, a lot still remains to be done as there is currently no overarching general foundation and methodology.

This seminar has been organized to bring together researchers from different communities, such as automated proving, interactive proving and SAT/SMT solving as well as from logic, proof engineering, program verification and formal mathematics. An essential goal has been to form a community around these issues in order to learn about and reconcile these different approaches. This will allow us to develop a common objective and framework for proof developments that support the communication, reuse, and interoperability of proofs.

The program of the seminar included introductions to different methods and techniques, the definition of precise objectives, and the description of recent achievements and current trends. It consisted of 30 contributed talks from experts on the above topics and six breakout sessions on major problems: theory graph – based reasoning, benchmarks, conflicting logics and system designs, proof certificates, design of a universal library of elementary mathematics, and a standard for system integration and proof interchange. The contributed talks took place in the morning, and two parallel breakout sessions each took place on Monday, Tuesday and Thursday afternoon, followed by plenary discussions organized by each session’s moderator.

The organizers would like to thank the Dagstuhl team and all the participants for making this first seminar a success and, hopefully, an event to be repeated.

2 Table of Contents

Executive Summary

Gilles Dowek, Catherine Dubois, Brigitte Pientka, and Florian Rabe 75

Overview of Talks

Translating between Agda and Dedukti
Andreas Martin Abel 79

Transferring Lemmas and Proofs in Isabelle/HOL: a Survey
Jesús María Aransay Azofra 79

Uniform Proofs via Shallow Semantic Embeddings?
Christoph Benz Müller 79

Are Translations between Proof Assistants Possible or Even Desirable at All?
Jasmin Christian Blanchette 80

Using External Provers in Proof Assistants
Frédéric Blanqui 80

The Continuity of Monadic Stream Functions
Venanzio Capretta 81

Reengineering Proofs in Dedukti: an Example
Gilles Dowek 82

FoCaLiZe and Dedukti to the Rescue for Proof Interoperability
Catherine Dubois 82

We Need a Better Style of Proof
William M. Farmer 82

Comparing Systems for Reasoning with Higher-Order Abstract Syntax Representations
Amy Felty 83

Aligning Concepts across Proof Assistant Libraries
Thibault Gauthier 83

Extending Higher-order Logic with Predicate Subtyping
Frédéric Gilbert 83

Inference Systems for Satisfiability Problems
Stéphane Graham-Lengrand 84

Not Incompatible Logics
Olivier Hermant 85

Lazy Proofs for DPLL(T)-Based SMT Solvers
Guy Katz 85

The Triumvirate of Automation, Expressivity, and Safety
Chantal Keller 85

Reproducibility, Trust, and Proof Checkings
Dale Miller 86

Benchmarks for Mechanized Meta-theory: a very Personal and Partial View
Alberto Momigliano 86

| | |
|---|----|
| Mechanizing Meta-Theory in Beluga <i>Brigitte Pientka</i> | 87 |
| On Universality of Proof Systems <i>Elaine Pimentel</i> | 87 |
| MMT: A UniFormal Approach to Knowledge Representation <i>Florian Rabe</i> | 88 |
| Higher Order Constraint Logic Programming for Interactive Theorem Proving <i>Claudio Sacerdoti Coen</i> | 88 |
| LLFP: a Framework for Interconnecting Logical Frameworks <i>Ivan Scagnetto</i> | 88 |
| External termination proofs for Isabelle with IsaFoR and CeTA <i>René Thiemann</i> | 89 |
| Parsing Mathematics by Learning from Aligned Corpora and Theorem Proving <i>Josef Urban</i> | 89 |
| Plugging External Provers into the Rodin Platform <i>Laurent Voisin</i> | 90 |
| Computation in Proofs <i>Freek Wiedijk</i> | 90 |
| Ancient History of the Quest for Universality of Proofs <i>Bruno Woltzenlogel Paleo</i> | 91 |
| First-Order Conflict-Driven Clause Learning from a Proof-Theoretical Perspective <i>Bruno Woltzenlogel Paleo</i> | 91 |
| Working groups | |
| Breakout Session on Theory Graph Based Reasoning <i>William M. Farmer</i> | 92 |
| Breakout Session on Conflicting Logics and System Designs <i>Olivier Hermant and Chantal Keller</i> | 92 |
| Breakout Session on a Universal Library <i>Michael Kohlhase and Catherine Dubois</i> | 93 |
| Breakout session on A standard for system integration and proof interchange <i>Ramana Kumar and Florian Rabe</i> | 94 |
| Breakout Session on Proof Certificates <i>Dale Miller</i> | 95 |
| Breakout Session on Benchmarks <i>Alberto Momigliano and Amy Felty</i> | 96 |
| Participants | 98 |

3 Overview of Talks

3.1 Translating between Agda and Dedukti

Andreas Martin Abel (Chalmers UT – Göteborg, SE)

License © Creative Commons BY 3.0 Unported license
© Andreas Martin Abel

Boespflug and Burel have designed CoqInE, a translation of Coq’s Calculus of Inductive Constructions into Dedukti, the logical framework with rewriting. We conjecture a similar translation could be possible from Agda to Dedukti. The reverse translation is possible since Agda recently got extended with rewrite tools. In this talk, we propose an implementation of translations between Agda and Dedukti.

3.2 Transferring Lemmas and Proofs in Isabelle/HOL: a Survey

Jesús María Aransay Azofra (University of La Rioja – Logroño, ES)

License © Creative Commons BY 3.0 Unported license
© Jesús María Aransay Azofra

Data types admit different representations. The reasons to prioritise one or another range from efficiency to simplicity. Interactive theorem provers allow users to work with these representations and offer tools that ease their communication. These tools are required to preserve the formalism among the representations. In this talk we present some use cases in the proof assistant Isabelle/HOL, defining the scope of each of these tools and their possible scenarios.

3.3 Uniform Proofs via Shallow Semantic Embeddings?

Christoph Benz Müller (FU Berlin, DE)

License © Creative Commons BY 3.0 Unported license
© Christoph Benz Müller

Main reference C. Benz Müller, B. Woltzenlogel Paleo, “The Inconsistency in Gödel’s Ontological Argument: A Success Story for AI in Metaphysics”, IJCAI 2016, pp. 936–942, 2016.

URL <http://www.ijcai.org/Abstract/16/137>

Many classical and non-classical logics can be elegantly mechanised and automated by exploiting shallow semantical embeddings in classical higher-order logic. In recent research this approach has been successfully applied in various disciplines, including metaphysics, mathematics and artificial intelligence. Moreover, it has recently been utilised as the core framework in my (awarded) lecture course on Computational Metaphysics at Freie Universität Berlin. In this talk I demonstrate the approach and discuss its potential for achieving uniform proofs across various logics.

References

- 1 Christoph Benz Müller and Dana Scott, *Axiomatizing Category Theory in Free Logic*. arXiv, <http://arxiv.org/abs/1609.01493>, 2016.

- 2 Christoph Benzmüller and Dana Scott, *Automating Free Logic in Isabelle/HOL*. In *Mathematical Software – ICMS 2016, 5th International Congress, Proceedings* (G.-M. Greuel, T. Koch, P. Paule, A. Sommese, eds.), Springer, LNCS, volume 9725, pp. 43-50, 2016.
- 3 Christoph Benzmüller, Max Wisniewski and Alexander Steen, *Computational Metaphysics – Bewerbung zum zentralen Lehrpreis der Freien Universität Berlin, FU Berlin*, 2015.
- 4 Christoph Benzmüller and Bruno Woltzenlogel Paleo, *The Inconsistency in Gödel’s Ontological Argument: A Success Story for AI in Metaphysics*. In *IJCAI 2016* (Subbarao Kambhampati, ed.), AAAI Press, volume 1-3, pp. 936-942, 2016.
- 5 Christoph Benzmüller and Bruno Woltzenlogel Paleo, *Automating Gödel’s Ontological Proof of God’s Existence with Higher-order Automated Theorem Provers*. In *ECAI 2014* (Torsten Schaub, Gerhard Friedrich, Barry O’Sullivan, eds.), IOS Press, *Frontiers in Artificial Intelligence and Applications*, volume 263, pp. 93 – 98, 2014.

3.4 Are Translations between Proof Assistants Possible or Even Desirable at All?

Jasmin Christian Blanchette (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license

© Jasmin Christian Blanchette

Joint work of Jasmin Christian Blanchette, Sascha Böhme, Mathias Fleury, Steffen Juilf Smolka, Albert Steckermeier

Main reference J. C. Blanchette, S. Böhme, M. Fleury, S. J. Smolka, A. Steckermeier, “Semi-intelligible Isar proofs from machine-generated proofs”, *J. Autom. Reasoning*, Vol. 56(2), pp. 155–200, Springer, 2016.

URL <https://dx.doi.org/10.1007/s10817-015-9335-3>

Combining proofs developed using different proof assistants would seem to be highly desirable. After all, a lot of effort goes into formalizing a result in one assistant, and it makes sense to reuse it as much as possible. However, there are lots of obstacles before we have tools that can be widely deployed. When Isabelle/HOL users port analysis results from HOL Light, in practice they cannot rely on automatic bridges such as OpenTheory, even though both systems are based on simple type theory. I will review different approaches to prove exchange and emphasize their shortcomings from the viewpoint of end users. I will also briefly describe my work on translation of proofs between automatic theorem provers and Isabelle/HOL.

3.5 Using External Provers in Proof Assistants

Frédéric Blanqui (ENS – Cachan, FR)

License © Creative Commons BY 3.0 Unported license

© Frédéric Blanqui

URL <http://cl-informatik.uibk.ac.at/software/cpf/>

Using external provers in proof assistants is an example of small scale inter-operability. It relieves proof assistant users and proof assistant developers. The main obstacle is to be able to certify the results of the external prover. It is now well established for proving propositional or first-order subgoals, except perhaps in proof assistants using dependent types. But it is also possible to use external provers for termination and confluence problems. Indeed, since 2009, there is a common format called CPF for termination certificates and certified tools to check their correctness.

3.6 The Continuity of Monadic Stream Functions

Venanzio Capretta (*University of Nottingham, GB*)

License  Creative Commons BY 3.0 Unported license
© Venanzio Capretta

Streams are infinite sequences of values. They inhabit a frontier region of constructive mathematics and computer science: they cannot be represented fully inside human minds and computer memories, but they are omnipresent as input, output and interactive behaviour.

Brouwer formulated the notion of “choice sequence”, a progression of values that is not generated by an effective rule, but rather by a creative subject. Alternatively, they may model repeated measurement of physical phenomena or interactive input from a non-predictable user.

If the streams themselves are not computable, functions on them, realized as programs, must be effective. From this requirement, Brouwer concluded that a Continuity Principle must hold: all functions on streams of natural numbers are continuous. This means that the value of a function on a specific stream only depends on a finite initial segment.

The principle seems to be justifiable in a computational view and is certainly true from the meta-theoretical standpoint. We may be tempted to add it in a formulation of the foundations of constructive mathematics. However, recently Martín Escardó discovered that if we add the Continuity Principle to Constructive Type Theory, we obtain a contradiction. I will discuss the paradox and the possible avenues of repair. One way is to weaken the principle, using an existential quantifier that does not provide a witness.

I suggest a different solution. In the original formulation, we consider functions on the internal type of streams, encoded as functions from natural numbers to natural numbers. But this type does not capture the idea of an unpredictable sequence not subject to rule and possibly coming from an outside source. I propose that “monadic streams” are a better model: these are sequences in which a monadic action must be executed to obtain the next element and the continuation. A monadic action is any of a wide class of enriched structures and modes of value presentation. Monadic programming has been very successful in modelling interaction and side effects in functional programming.

I propose a version of the Continuity Principle for monadic streams. This has great potential not only as a foundational theory but in practical applications. Monadic streams have already been successfully used in functional reactive programming and game implementation. The principle applies to functions on monadic streams that are polymorphic in the monad and natural on it. They are a reasonable description of mappings on sequences that do not depend on how the sequence is generated. I will prove that these functions are always continuous.

I think these issues are extremely important for the future of computer-assisted mathematics. Monadic streams are a very promising data structure, needed to model reactive continuous processes. This work shows that they are also relevant in the design of the logical principles underlying formalized mathematics.

3.7 Reengineering Proofs in Dedukti: an Example

Gilles Dowek (INRIA & ENS Cachan, FR)

License  Creative Commons BY 3.0 Unported license
© Gilles Dowek

The system Dedukti is a logical framework. We illustrate how it can be used to reengineer proofs with the example of the translation to Simple type theory of proofs expressed in the Calculus of constructions.

3.8 FoCaLiZe and Dedukti to the Rescue for Proof Interoperability

Catherine Dubois (ENSIIE – Evry, FR)

License  Creative Commons BY 3.0 Unported license
© Catherine Dubois
Joint work of Raphaël Cauderlier, Catherine Dubois

We propose a methodology to combine proofs coming from different provers relying on Dedukti as a common formalism in which proofs can be translated and combined. To relate the independently developed mathematical libraries used in proof assistants, we rely on the structuring features offered by FoCaLiZe. We illustrate this methodology on the Sieve of Eratosthenes, which we prove correct using HOL and Coq in combination.

3.9 We Need a Better Style of Proof

William M. Farmer (McMaster University – Hamilton, CA)

License  Creative Commons BY 3.0 Unported license
© William M. Farmer

Proofs serve several diverse purposes in mathematics. They are used to communicate mathematical ideas, certify that mathematical results are correct, discover new mathematical facts, learn mathematics, establish the interconnections between mathematical ideas, show the universality of mathematical results, and create mathematical beauty. Traditional proofs and (computer-supported) formal proofs do not fulfill these purposes equally well. In fact, traditional proofs serve some purposes much better than formal proofs, and vice versa. For example, traditional proofs are usually better for communication, while formal proofs are usually better for certification. We compare both traditional and formal proofs with respect to these seven purposes and show that both styles of proof have serious shortcomings. We offer a new style of proof in which (1) informal and formal proof components are combined (in accordance with Michael Kohlhase’s notion of flexiformality), (2) results are proved at the optimal level of abstraction (in accordance with the little theories method), and (3) cross-checks are employed systematically. We argue that this style of proof fulfills the purposes of mathematical proofs much better than both traditional and formal proofs.

3.10 Comparing Systems for Reasoning with Higher-Order Abstract Syntax Representations

Amy Felty (University of Ottawa, CA)

License © Creative Commons BY 3.0 Unported license
© Amy Felty

Joint work of Amy Felty, Alberto Momigliano, Brigitte Pientka

Over the past three decades, a variety of meta-reasoning systems which support reasoning about higher-order abstract specifications have been designed and developed. We summarize our work on surveying and comparing four meta-reasoning systems, Twelf, Beluga, Abella and Hybrid, using several benchmarks from the open repository ORBI that describes challenge problems for reasoning with higher-order abstract syntax representations. In particular, we investigate how these systems mechanize and support reasoning using a context of assumptions. This highlights commonalities and differences in these systems and is a first step towards translating between them.

3.11 Aligning Concepts across Proof Assistant Libraries

Thibault Gauthier (Universität Innsbruck, AT)

License © Creative Commons BY 3.0 Unported license
© Thibault Gauthier

Joint work of Thibault Gauthier, Cezary Kaliszyk

As the knowledge available in the computer understandable proof corpora grows, recognizing repeating patterns becomes a necessary requirement in order to organize, synthesize, share, and transmit ideas. In this work, we automatically discover patterns in the libraries of interactive theorem provers and thus provide the basis for such applications for proof assistants. This involves detecting close properties, inducing the presence of matching concepts, as well as dynamically evaluating the quality of matches from the similarity of the environment of each concept. We further propose a classification process, which involves a disambiguation mechanism to decide which concepts actually represent the same mathematical ideas. We evaluate the approach on the libraries of six proof assistants based on different logical foundations: HOL4, HOL Light, and Isabelle/HOL for higher-order logic, Coq and Matita for intuitionistic type theory, and the Mizar Mathematical Library for set theory. Comparing the structures available in these libraries our algorithm automatically discovers hundreds of isomorphic concepts and thousands of highly similar ones.

3.12 Extending Higher-order Logic with Predicate Subtyping

Frédéric Gilbert (ENS - Cachan, FR)

License © Creative Commons BY 3.0 Unported license
© Frédéric Gilbert

Predicate subtyping is an extension of higher-order logic where the grammar of types is enriched with a construction for restricted comprehension allowing to define, for any type A and any predicate P on A , a new type $\{A \mid P\}$. The inhabitants of such a type are the inhabitants t of A for which $P(t)$ is provable. As a consequence, type-checking becomes

undecidable. We present a possible formalization of predicate subtyping, which can be the base of a formalization of the proof assistant PVS. We also present a similar system using explicit proofs and coercions. This system is used as a lightweight language for predicate subtyping. It is also a first step towards the expression of predicate subtyping in a universal system.

3.13 Inference Systems for Satisfiability Problems

Stéphane Graham-Lengrand (Ecole Polytechnique – Palaiseau, FR)

License © Creative Commons BY 3.0 Unported license
© Stéphane Graham-Lengrand

Joint work of Maria Paola Bonacina, Stéphane Graham-Lengrand, Natarajan Shankar

One of the most popular approaches for solving propositional SAT problems is CDCL (Conflict-Driven Clause Learning), a variant of DPLL where model construction steps alternate with conflict analysis steps. In terms of proof theory, this is an alternation between bottom-up and top-down applications of rules from an inference system.

MCSat is a methodology for generalising CDCL to other theories than propositional logic. It thereby addresses (quantifier-free) SAT-Modulo-Theories problems, but in a way that seems rather different from the widely used architecture where DPLL interacts with the combination, by the Nelson-Oppen method, of theory-specific decision procedures.

We identify the notion of an MCSat-friendly inference system, and define a generic MCSat calculus that is sound and complete for satisfiability in the union of n arbitrary theories (including for instance propositional logic), as long as each of them comes with an MCSat-friendly inference system.

We show how the Nelson-Oppen method is a particular case of our MCSat-combination method, reconciling the widely-implemented technique with the new MCSat ideas.

References

- 1 M. P. Bonacina, S. Graham-Lengrand, and N. Shankar. A model-constructing framework for theory combination. Research report, Università degli Studi di Verona – SRI International – CNRS – INRIA, 2016. Available at <http://hal.archives-ouvertes.fr/hal-01425305>
- 2 Leonardo de Moura and Dejan Jovanović. A model-constructing satisfiability calculus. In Roberto Giacobazzi, Josh Berdine, and Isabella Mastroeni, editors, *Proc. of the 14th Int. Conf. on Verification, Model Checking and Abstract Interpretation (VMCAI)*, volume 7737 of *LNCS*, pages 1–12. Springer, 2013.
- 3 Dejan Jovanović, Clark Barrett, and Leonardo de Moura. The design and implementation of the model-constructing satisfiability calculus. In Barbara Jobstman and Sandip Ray, editors, *Proc. of the 13th Conf. on Formal Methods in Computer Aided Design (FMCAD)*. ACM and IEEE, 2013.

3.14 Not Incompatible Logics

Olivier Hermant (Ecole des Mines de Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Olivier Hermant

The formalisms used to express proofs are most of the time incompatible, either in a strong form, inconsistent with each other, or in a weaker form, leading to different properties of the logic.

In this talk, I introduce the story of overcoming this difficulty on the example of classical and intuitionistic logic.

3.15 Lazy Proofs for DPLL(T)-Based SMT Solvers

Guy Katz (Stanford University, US)

License © Creative Commons BY 3.0 Unported license
© Guy Katz

Joint work of Guy Katz, Clark Barrett, Cesare Tinelli, Andrew Reynolds, Liana Hadarean

Main reference G. Katz, C. Barrett, C. Tinelli, A. Reynolds, L. Hadarean, “Lazy Proofs for DPLL(T)-Based SMT Solvers”, in Proc. of the 16th Int’l Conf. on Formal Methods in Computer-Aided Design (FMCAD), pp. 93–100, 2016.

URL <https://stanford.edu/~guyk/pub/FMCAD2016.pdf>

With the integration of SMT solvers into analysis frameworks aimed at ensuring a system’s end-to-end correctness, having a high level of confidence in these solvers’ results has become crucial. For unsatisfiable queries, a reasonable approach is to have the solver return an independently checkable proof of unsatisfiability. We propose a lazy, extensible and robust method for enhancing DPLL(T)-style SMT solvers with proof-generation capabilities. Our method maintains separate Boolean-level and theory-level proofs, and weaves them together into one coherent artifact. Each theory-specific solver is called upon lazily, a posteriori, to prove precisely those solution steps it is responsible for and that are needed for the final proof. We present an implementation of our technique in the CVC4 SMT solver, capable of producing unsatisfiability proofs for quantifier-free queries involving uninterpreted functions, arrays, bitvectors and combinations thereof. We discuss an evaluation of our tool using industrial benchmarks and benchmarks from the SMTLIB library, which shows promising results.

3.16 The Triumvirate of Automation, Expressivity, and Safety

Chantal Keller (University of Paris Sud – Orsay, FR)

License © Creative Commons BY 3.0 Unported license
© Chantal Keller

In this survey, I will analyze various approaches to interoperability between proof systems: what effort does this interoperability requires? Can two systems be really agnostic of each other to communicate? How deep can we go into automation, expressivity and safety?

In particular, I will present:

- the interoperability a posteriori between already established interactive and automatic theorem provers, such as SMTCoq or Ergo for the Coq proof assistant or sledgehammer for the Isabelle/HOL proof assistant;
- the interoperability a priori inside proof systems that are designed to be automatic, expressive and safe in a more tightened way, such as lean, F* or Why3.

3.17 Reproducibility, Trust, and Proof Checkings

Dale Miller (INRIA Saclay – Île-de-France, FR)

License  Creative Commons BY 3.0 Unported license
© Dale Miller

Formal proofs are produced and checked by machines. Machines are physical devices, of course, and their software and their execution are subject to errors. As in other scientific domains, reproducibility is key to establishing trust, whether it is a claim in physics or a claim that a given file contains a valid proof. A high degree of trust in a formal proof comes from executing a trusted proof checker on a claimed proof, thereby, reproducing the claim. In order to trust a proof checker, it should be possible to implement new proof checkers or to exam the source of existing provers and to be convinced that they are sound implementations of logic. Providing a formal semantics for proof languages is an important step in allowing for this kind of independent and trustworthy proof checking to be achieved.

3.18 Benchmarks for Mechanized Meta-theory: a very Personal and Partial View

Alberto Momigliano (University of Milan, IT)

License  Creative Commons BY 3.0 Unported license
© Alberto Momigliano

Joint work of Amy Felty, Alberto Momigliano, Brigitte Pientka

Benchmarks in theorem proving have been very useful, made the state of the art progress or at least take stock, as the bright example of *TPTP* testifies, whose influence on the development, testing and evaluation of automated theorem provers cannot be underestimated. The situation is less satisfactory for proof assistants, where each system comes with its own set of examples/libraries, some of them gigantic. This is not surprising, since we are potentially addressing the whole realm of mathematics.

In this talk I try to evaluate the impact, if any, that benchmarks have had on the sub-field of the meta-theory of deductive systems, such as the ones studied in Programming Language Theory, and its feedback, again if any, on the development of logical frameworks.

References

- 1 Amy P. Felty, Alberto Momigliano, Brigitte Pientka: *The Next 700 Challenge Problems for Reasoning with Higher-Order Abstract Syntax Representations – Part 2 – A Survey*. J. Autom. Reasoning 55(4): 307-372 (2015)
- 2 Amy P. Felty, Alberto Momigliano, Brigitte Pientka: *An Open Challenge Problem Repository for Systems Supporting Binders*. LFMTP 2015: 18-32

3.19 Mechanizing Meta-Theory in Beluga

Brigitte Pientka (McGill University – Montreal, CA)

License © Creative Commons BY 3.0 Unported license
© Brigitte Pientka

Joint work of Andrew Cave, Brigitte Pientka

Mechanizing formal systems, given via axioms and inference rules, together with proofs about them plays an important role in establishing trust in formal developments. In this talk, I will survey the proof environment Beluga. To specify formal systems and represent derivations within them, Beluga provides a sophisticated infrastructure based on the logical framework LF; to reason about formal systems, Beluga provides a dependently typed functional language for implementing inductive proofs about derivation trees as recursive functions following the Curry-Howard isomorphism. Key to this approach is the ability to model derivation trees that depend on a context of assumptions using a generalization of the logical framework LF, i.e. contextual LF which supports first-class contexts and simultaneous substitutions.

Our experience has demonstrated that Beluga enables direct and compact mechanizations of the meta-theory of formal systems, in particular programming languages and logics. To demonstrate Beluga’s strength in this talk, we develop a weak normalization proof using logical relations.

References

- 1 A. Cave and B. Pientka. *Programming with binders and indexed data-types*. In *POPL’12*, pages 413–424. ACM, 2012.
- 2 A. Cave and B. Pientka. *A case study on logical relations using contextual types*. In *LFMTP’15*, pages 18–33. Electr. Proc. in Theoretical Computer Science (EPTCS), 2015.
- 3 B. Pientka. *A type-theoretic foundation for programming with higher-order abstract syntax and first-class substitutions*. In *POPL’08*, pages 371–382. ACM, 2008.
- 4 B. Pientka and A. Cave. *Inductive Beluga: Programming proofs (system description)*. In *CADE-25*, LNCS 9195, pages 272–281. Springer, 2015.

3.20 On Universality of Proof Systems

Elaine Pimentel (Federal University of Rio Grande do Norte, BR)

License © Creative Commons BY 3.0 Unported license
© Elaine Pimentel

Joint work of Björn Lellmann, Carlos Olarte, Elaine Pimentel

We propose a notion of modular linear nested sequent calculi (LNS) for different modalities which brings down the complexity of proof search to that of the corresponding sequent calculi. Examples include normal and non-normal classical modal logics as well as multiplicative additive linear logic (MALL) plus simply dependent multimodalities. Since LNS systems can be adequately encoded into (plain) linear logic, LL can be seen, in fact, as an “universal framework” for the specification of logical systems. While the modularity of the systems lead to a generic way of building theorem provers for different logics (all of them based on the same grounds), universality of LL allows for the use of the same logical framework for reasoning about all such logical systems.

3.21 MMT: A UniFormal Approach to Knowledge Representation

Florian Rabe (Jacobs University Bremen, DE)

License  Creative Commons BY 3.0 Unported license
 © Florian Rabe
URL <http://uniformal.github.io/>

UniFormal is the idea of representing all aspects of knowledge uniformly, including narration, deduction, computation, and databases. Moreover, it means to abstract from the multitude of individual systems, which not only often focus on just one aspect but are doing so in mutually incompatible ways, thus creating a universal framework of formal knowledge.

MMT is a concrete representation language to that end. It systematically abstracts from assumptions typically inherent in the syntax and semantics of concrete systems, and focuses on language-independence, modularity, and system interoperability. While constantly evolving in order to converge towards UniFormal, its design and implementation have become very mature. It is now a readily usable high-level platform for the design, analysis, and implementation of formal systems.

This talk gives an overview of the current state of MMT, its existing successes and its future challenges.

3.22 Higher Order Constraint Logic Programming for Interactive Theorem Proving

Claudio Sacerdoti Coen (University of Bologna, IT)

License  Creative Commons BY 3.0 Unported license
 © Claudio Sacerdoti Coen

Some interactive theorem provers, like Coq, Matita and Agda are implemented around the Curry-Howard isomorphism. Proof checking is type checking, and it can be compactly represented in an Higher Order Logic Programming (HOLP) language / logical framework. Interactive proof construction, however, requires the manipulation of terms containing metavariables, and a significant amount of logic independent code to accommodate metavariables (and narrowing) in “type checking” (aka elaboration, refinement). We propose to delegate such work to the metalanguage by extending HOLP with Constraint Programming features induced by a delay mechanism for “too flexible” goals.

3.23 LLFP: a Framework for Interconnecting Logical Frameworks

Ivan Scagnetto (University of Udine, IT)

License  Creative Commons BY 3.0 Unported license
 © Ivan Scagnetto

Joint work of F. Honsell, M. Lenisa, L. Liquori, P. Maksimovic, V. Michielini, Ivan Scagnetto
Main reference F. Honsell, L. Liquori, P. Maksimovic, I. Scagnetto, “LLFP: A Logical Framework for Modeling External Evidence, Side Conditions, and Proof Irrelevance using Monads”, 2016.
URL https://users.dimi.uniud.it/~ivan.scagnetto/LLFP_LMCS.pdf

LLFP (Lax LF with Predicates) is an extension of Edinburgh Logical Framework (LF) with locking type constructors and with a family of monads indexed by predicates over typed terms. Locks are a sort of modality constructors, releasing their argument under the condition that

a predicate, possibly external to the system, is satisfied on an appropriate typed judgement. This mechanism paves the way for a proof assistant allowing the user to make calls to external oracles (e.g., other proof assistants) during the proof development activity. Such calls are usually made to factor out the complexity of encoding specific features of logical systems which would otherwise be awkwardly encoded in LF, e.g. side-conditions in the application of rules in Modal Logics, and sub-structural rules, as in noncommutative Linear Logic. Using LLFP, these conditions need only to be specified, while their verification can be delegated to an external proof engine, according to the Poincaré Principle. Moreover, monads also express the effect of postponing verifications. This fact allows the user to focus on the main proof, leaving the possibly external verification of details at the end. A first prototype of a type checker for LLFP has been recently written in OCaml by V. Michielini (ENS Lyon): the software currently supports the Coq System as an external proof assistant.

3.24 External termination proofs for Isabelle with IsaFoR and CeTA

René Thiemann (Universität Innsbruck, AT)

License © Creative Commons BY 3.0 Unported license
© René Thiemann

Joint work of Alexander Krauss, Christian Sternagel, René Thiemann, Carsten Fuhs, Jürgen Giesl

Main reference A. Krauss, C. Sternagel, R. Thiemann, C. Fuhs, J. Giesl, “Termination of Isabelle Functions via Termination of Rewriting”, in Proc. of the 2nd Int’l Conf. on Interactive Theorem Proving (ITP’11), LNCS Vol. 6898, pp. 152–167, Springer, 2011.

URL http://dx.doi.org/10.1007/978-3-642-22863-6_13

CeTA is a certifier for automatically generated termination proofs, which supports a wide variety of termination techniques. Its soundness is proven in IsaFoR, the Isabelle formalization of rewriting.

We will present an overview of the capabilities of CeTA, and also discuss to which extent CeTA can be used to discharge termination proof obligations that arise from function definitions in Isabelle itself.

3.25 Parsing Mathematics by Learning from Aligned Corpora and Theorem Proving

Josef Urban (Czech Technical University – Prague, CZ)

License © Creative Commons BY 3.0 Unported license
© Josef Urban

Joint work of Cezary Kaliszyk, Josef Urban, Jiri Vyskocil

Main reference C. Kaliszyk, J. Urban, J. Vyskocil, “Learning to Parse on Aligned Corpora (Rough Diamond)”, in Proc. of the Int’l Conf. on Interactive Theorem Proving (ITP 2015), LNCS, Vol. 9236, pp. 227–233, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-22102-1_15

One of the biggest hurdles that mathematicians encounter when working with formal proof assistants is the necessity to get acquainted with the formal terminology and the parsing mechanisms used in formal proof. While overloading and syntactic ambiguity are ubiquitous in regular mathematics, theorem proving requires full formality. This makes computer verification of mathematical proofs a laborious and so far rare enterprise. In this work we start to address this problem by developing probabilistic AI methods that autonomously train disambiguation on large aligned corpora of informal and formal mathematical formulas.

The resulting parse trees are then filtered by strong semantic AI methods such as large-theory automated theorem proving. We describe the general motivation and our first experiments, and show an online system for parsing ambiguous formulas over the Flyspeck library.

References

- 1 Blanchette, J. C.; Kaliszyk, C.; Paulson, L. C.; and Urban, J. *Hammering towards QED*. J. Formalized Reasoning. 9(1):101–148. 2016.
- 2 Kaliszyk, C., and Urban, J. *Learning-assisted automated reasoning with Flyspeck*. J. Autom. Reasoning 53(2):173–213. 2014.
- 3 Kaliszyk, C.; Urban, J.; and Vyskocil, J. *Learning to parse on aligned corpora (rough diamond)*. In Urban, C., and Zhang, X., eds., Interactive Theorem Proving - 6th International Conference, ITP 2015, Nanjing, China, August 24-27, 2015, Proceedings, volume 9236 of *Lecture Notes in Computer Science*, 227–233. Springer.
- 4 Tankink, C.; Kaliszyk, C.; Urban, J.; and Geuvers, H. *Formal mathematics on display: A wiki for Flyspeck*. In Carette, J.; Aspinall, D.; Lange, C.; Sojka, P.; and Windsteiger, W., eds., *MKM/Calculemus/DML*, volume 7961 of *LNCS*, 152–167. Springer. 2013.
- 5 Zinn, C. *Understanding informal mathematical discourse*. Ph.D. Dissertation, University of Erlangen-Nuremberg. 2004.

3.26 Plugging External Provers into the Rodin Platform

Laurent Voisin (SYSTEREL Aix-en-Provence, FR)

License © Creative Commons BY 3.0 Unported license
© Laurent Voisin

Main reference D. Déharbe, P. Fontaine, Y. Guyot, L. Voisin, “Integrating SMT solvers in Rodin”, *Science of Computer Programming*, Vol. 94, pp. 130–143, Elsevier, 2014.

URL <http://dx.doi.org/10.1016/j.scico.2014.04.012>

The Rodin platform allows to model reactive systems and prove them correct using the Event-B formal notation. The mathematical logic used in classical first-order predicate calculus with equality, set theory and integer arithmetic. The proof are expressed in the sequent calculus, where the inference rules are computed by external reasoners. Some reasoners are implemented by connecting external provers, which provides terminating inference rules. These external provers allow to reduce drastically the need to perform manual proofs, by providing the automation to discharge all trivial facts.

3.27 Computation in Proofs

Freek Wiedijk (Radboud University Nijmegen, NL)

License © Creative Commons BY 3.0 Unported license
© Freek Wiedijk

I discuss the Poincare Principle: the notion that calculations do not need to be proved. As part of this I show a small experiment to add a Poincare Principle to HOL Light.

3.28 Ancient History of the Quest for Universality of Proofs

Bruno Woltzenlogel Paleo (Australian National University – Canberra, AU)

License  Creative Commons BY 3.0 Unported license
© Bruno Woltzenlogel Paleo

This was the second last talk in a seminar where much had already been said about the present and future of universality of proofs. To complement that, I decided to talk briefly about the distant past, sharing interesting facts about Leibniz, which I learned during a historical research triggered by the 300th anniversary of his death. The talk was based on an analysis of selected quotations from Leibniz, which give insight into what Leibniz would have thought if he could see today's state of the art.

Leibniz was a pioneer in the topics of the seminar. Three and a half centuries ago he already dreamt of a universal logical language (*characteristica universalis*) and a reasoning calculus. But his contribution was not only a dream. He also took concrete initial steps to fulfil his dream, by defining his own language for an algebra of concepts and even describing how to encode its logical sentences into arithmetical expressions that automated calculating machines of his time could handle. While Leibniz desired a universal logical language because he had none, today we seek universality because of we have too many logics and proof languages competing for acceptance. This is a clear, sign of the astonishing success achieved by our community so far. Although somewhat ironic, the plurality of alternatives is a good problem to have.

The potential of a universal logic for solving concrete controversies among people was a major motivation for Leibniz, who also explicitly aimed at all fields of inquiry capable of certainty. When he compares mathematics and metaphysics, for instance, Leibniz shows that he considered mathematics neither controversial enough nor in particular need of extremely precise formal reasoning. In contrast, today's applications of automated reasoning are still heavily biased towards mathematics. Despite a few exceptions, the mainstream attitude is currently not yet as universal with respect to application domains as it could be.

Leibniz was also overly optimistic about how easy it would be to learn a universal logical language. He wanted it to be so simple that anyone could learn it in a week or two. But the most sophisticated expressive universal languages that we have today may still require semester-long advanced courses for gifted students who already have a strong background in logic. Nevertheless, user interfaces for theorem provers have been progressing rapidly and maybe it will not take long for our technology to become universally accessible to all after only a short period of training.

3.29 First-Order Conflict-Driven Clause Learning from a Proof-Theoretical Perspective

Bruno Woltzenlogel Paleo (Australian National University – Canberra, AU)

License  Creative Commons BY 3.0 Unported license
© Bruno Woltzenlogel Paleo

In this talk I present the new (first-order) conflict resolution calculus: an extension of the resolution calculus inspired by techniques used in modern SAT-solvers. The resolution inference rule is restricted to (first-order) unit propagation and the calculus is extended with a mechanism for assuming decision literals and with a new inference rule for clause

learning, which is a first-order generalization of the propositional conflict-driven clause learning (CDCL) procedure. The calculus is sound (because it can be simulated by natural deduction) and refutationally complete (because it can simulate resolution).

4 Working groups

4.1 Breakout Session on Theory Graph Based Reasoning

William M. Farmer (McMaster University – Hamilton, CA)

License  Creative Commons BY 3.0 Unported license
© William M. Farmer

A *theory graph* is a network of *axiomatic theories* linked by *meaning-preserving mappings*. The theories serve as abstract mathematical models and the mappings serve as information conduits that enable definitions and theorems to be passed from an abstract setting to many other usually more concrete settings. The theories may have different underlying *logics* and *foundations*.

In the first part of the session, we discussed how theory graphs can be used to represent mathematical knowledge and facilitate reasoning in a proof assistant. In the second part, we discussed the following questions:

1. What are examples of contemporary proof assistants and formal software specification systems that implement theory graph techniques?
2. What kind of objects and information (such as decision procedures and parsing/printing rules) can be attached to the theories in a theory graph?
3. What kind of reasoning is needed to build and exploit theory graphs?
4. How can theory graph technology be added to contemporary proof assistants in which all mathematical knowledge resides in a single theory?
5. Would the development of a logic-independent theorem prover for a system supporting theory graphs like MMT be a worthwhile project?

We were not able to achieve a consensus on what should be the answers to these questions. However, we did agree on the following action item: Select a set of theory graph techniques and compare how these techniques are implemented (if at all) in the leading proof assistants and formal software specification systems.

4.2 Breakout Session on Conflicting Logics and System Designs

Olivier Hermant (Ecole des Mines de Paris, FR) and Chantal Keller (University of Paris Sud – Orsay, FR)

License  Creative Commons BY 3.0 Unported license
© Olivier Hermant and Chantal Keller

We can observe many conflicts between logics, formal systems and even libraries inside the same tools. This session discussed in particular the following questions: how to take advantage in one system of the work in another system, the essence of conflicts in logic, and the ability to switch logics during a formalization process. A wide range of issues was tackled, forming a continuum between the research topics identified above. Some conflicts,

like set representation, can be resolved by defining morphisms, and coupling these with an abstraction step may ease the reuse of libraries across systems. This has led to several inter-platform developments, and stressed the need for a language that allows us to navigate between various levels of abstraction. Stronger conflicts, that lead to inconsistencies, might still be solved by a reverse analysis of proofs, so as to import only compatible, yet sufficient, slices of the frameworks, emphasizing the advantage to reason within little theories.

4.3 Breakout Session on a Universal Library

Michael Kohlhase (Universität Erlangen-Nürnberg, DE) and Catherine Dubois (ENSIIE – Evry, FR)

License © Creative Commons BY 3.0 Unported license
© Michael Kohlhase and Catherine Dubois

When formalizing mathematics, we usually need to rely on some knowledge which may or may not be formalized in proof assistants. Furthermore these theories may reside in many places and many forms. So the inventory of such formalized mathematical theories is not easy. The question raised in this breakout session concerns the requirements and design of a universal mathematical library. The discussion was organized following the Five W's method (When, What, Where, Who, Why?).

The first point discussed is the content of the universal library: participants agreed on limits, at least high school mathematics and wikipedia as the upper limit. Such a library should contain, for a notion or concept, definitions (multiple definitions if any), some examples and instances, its relations and dependencies, its main properties, a set of theorems characteristic to the properties and links to formal proofs. The next question is related to the organization of such data. Two directions were proposed: a glossary/dictionary or a theory graph. A first step would focus on the definitions of the concepts including their relations; a second step would be to collaborate on the data and develop some services. In this first step we can see many issues: how do we relate concepts? how do we take into account for parallel concepts? how can commutations be represented? what about the detection of problems? A possible solution is to rely on a graph of concepts where each node has a unique definition. Synonyms for similar concepts are attached to the node, giving different views. Examples could also be considered as views. A mechanism of composition is required. A quality control consisting in checking if concepts are similar is required.

Developing such a library is huge work. For example, wikipedia, PlanetMath, MathWorld count 100 000 concepts whereas a traditional mathematics dictionary has 35 000 words. Help should come from retired mathematicians, students, etc.

After having established these requirements, participants discussed the design of such a system. The conclusion was to build a prototype first, exploring existing systems, e.g. MathHub (<https://mathhub.info/>).

4.4 Breakout session on A standard for system integration and proof interchange

Ramana Kumar (Data61 / NICTA – Sydney, AU) and Florian Rabe (Jacobs University Bremen, DE)

License  Creative Commons BY 3.0 Unported license
© Ramana Kumar and Florian Rabe

Several requirements were put forward. Jasmin Blanchette suggested a standard similar to TSTP but with more structure, possibly a standardized subset of Isar. Freek Wiedijk suggested that the original source file of the proof should be recoverable and that the high-level structure of the proof should be apparent. Gilles Dowek pointed out that there are three categories of approach to this language: λ -terms, low-level proof steps (like LCF inference rules), or the statements of intermediate results (plus hints and other structure).

Other issues that were discussed were low level versus high level proofs, forwards versus backwards proofs and complete versus partial proofs, as well as the use of metadata to indicate the specific logic or its general features (e.g., being constructive).

The session then split into groups to gain more insight from considering concrete examples.

Finally the session discussed the following sketch by Florian Rabe for the core grammar of a possible proof standard:

| | | | |
|-----|-----|--|---|
| S | ::= | G is $C \vdash_{T^*}^L \{E\}$ by $\{P\}$ | theorem statement |
| P | ::= | E | proof term |
| | | $G(C, \{E\}^*, \{P\}^*)$ | operator/tactic applied to arguments |
| | | let C in $\{P\}$ | local definition |
| | | hence C by $P; \{P\}$ | forward step |
| | | goal C by $P; \{P\}$ | backward step |
| | | use E^* | partial proof |
| E | ::= | $G \mid X \mid \dots$ | expressions, terms, types, formulas, etc. |
| C | ::= | $(X [: E] [= E])^*$ | contexts |
| L | ::= | | logic identifier |
| T | ::= | | theory identifier |
| G | ::= | | global id from logic, theories, theorems |
| X | ::= | | local id introduced in proof |

Here curly brackets indicate the scope of the local identifiers introduced in the corresponding context, and $C \vdash_{T^*}^L E$ expresses the theorem “in logic L after importing the theories T^* we have for all C that E ”. A more refined version should include metadata to attach, e.g., original sources. It is straightforward to adapt the concrete syntax of existing standards such as TSTP or OMDoc to subsume (and possibly converge to) this abstract syntax.

To move forward, the session concluded that the community should collect standard prototypes and proof examples to better understand if this grammar suffices. Ramana Kumar and Florian Rabe volunteered to host this process using the repository <https://github.com/UniFormal/Proofs> which is open to and solicits community distributions.

4.5 Breakout Session on Proof Certificates

Dale Miller (INRIA Saclay – Île-de-France, FR)

License  Creative Commons BY 3.0 Unported license
© Dale Miller

Diversity

There is range of settings in which proofs and proof certificates are used. There are the familiar axis: classical vs intuitionistic and logic vs arithmetic¹. If we view formal proofs as a means to communicate between software systems, then such communication takes place across both time and space. If we examine short-distance and long-distance communication in these two dimensions, we have the following grid.

| Time | Space | Example of proof |
|-------|-------|---|
| Short | Short | A section in an interactive proof assistant can be dumped to disk in order to resume another day in the same proof assistant. |
| Long | Short | Completed proofs can be stored in a library associated with a particular proof assistant. |
| Short | Long | Cooperating but different provers may use specific certificates for their particular and immediate needs. |
| Long | Long | Proofs given high-level, declarative definitions may allow anyone to recheck them at anytime in the future. |

Another aspect of diversity occurs along the specific vs general spectrum. The current most significant use of proof certificates can be found in the areas where the role of logic is significantly constrained. For example, the following areas make use of well established proof formats: SAT solving (e.g., RUP, DRUP, DRAT), SMT (e.g. VeriT), and resolution refutation (e.g., IVY). Proof formats are also established for more encompassing logics: these include LF (λ II), LFSC, and TPTP. In the field of arithmetic, there are the OpenTheory project (for the HOL family of provers) and Dedukti (for the λ -cube). The Foundational Proof Certificate project is also attempting to find high-level definitions for a wide variety of proof certificates.

Insist on communication of proofs

There was a universal agreement in this breakout session that proof systems should provide options for outputting (exporting) proofs that they find. The following time-line was proposed in order to push the community towards the development of a standard format for proofs.

1. Provers need to be able to output some kind of useful information about their proofs. While the spirit of this proposal is meant to be informal, the intention is for developers of provers to make an effort to output documents that can be useful to others who want to consume proofs (whether to replay them in other systems, to extract information from them, etc). The goal is to insist on a commitment to an act of communication.

¹ We assumed that both induction and co-induction are treated within “arithmetic”.

2. The output from provers should be certified by proof checkers that are independent of the prover. Of course, there may be many different proof checkers which check proofs in a range of formats. The existence of independent proof checkers should start a move to the standardization of proof formats.
3. A single framework for certificates should be developed, tested, and analyzed. This effort builds on the previous two steps and is likely to contain both theoretical and engineering effort.
4. Develop a standard within some official standardization organization, such as ISO.

It is worth noting the role that competitions have played in helping to promote standards within the field. They provide a means for establishing an authority that is able to insist on standards.

What next?

While establishing a single standard for proof certificates seems to be at least several years away, it is worth noting that several hard problems remain even after we are able to make proofs and proof checking into a commodity. Since this breakout session was limited to the certification of proof, the following topics seem to be independent and not directly addressed.

1. Proofs generally are used for both certification and didactics. The problem of being able to read, browse, and interact with proofs and proof certificates was not addressed.
2. The existence of different theories for the same concepts (e.g., groups, real numbers, etc) was also not addressed. Generally theories are taken as axioms about various non-logical symbols and often there is no canonical selection of such non-logical symbols and their axiomization.

4.6 Breakout Session on Benchmarks

Alberto Momigliano (University of Milan, IT) and Amy Felty (University of Ottawa, CA)

License  Creative Commons BY 3.0 Unported license
© Alberto Momigliano and Amy Felty

This breakout session addressed the problem of designing benchmarks and challenge problems for interactive theorem proving systems. The discussion centered around four central questions, and resulted in partial answers and future directions of study. 1) The first question addressed why there is a need for benchmarks. One reason is to understand the differences between systems and highlight their strengths and limitations. Another reason is to use them as a starting point for the translation of theorem statements between systems. Furthermore, they can help to stimulate new development in the systems under study. 2) The group also addressed the question of whether we want universal benchmarks or different benchmarks in different areas. There was a general consensus for the latter. 3) The question of how formal benchmark descriptions should be is another important question. On one end of the spectrum, they could be informal natural language descriptions, and on the other end they could be formal parseable specifications. The informal text is always an important component, and there were varying degrees of support for more formal specifications. On the one hand, they provide a precise description, and on the other hand, formulating the theorem statement

precisely could be part of the challenge. 4) The last question addressed was the social process of developing benchmarks. We could appoint one or two people in the community to develop and collect benchmarks, and/or we could work on specific benchmarks in a few specific areas during the rest of the week. Two areas chosen for further discussion during the seminar were benchmarks involving binders (well-suited to systems supporting higher-order abstract syntax and related approaches) and benchmarks that involve induction/coinduction, fixpoints, corecursion, etc. (to test and better understand these capabilities in existing widely used proof assistants such as Coq, Isabelle, and Agda).

Participants

- Andreas Martin Abel
Chalmers UT – Göteborg, SE
- Jesús María Aransay Azofra
University of La Rioja –
Logroño, ES
- Christoph Benzmlüller
FU Berlin, DE
- Jasmin Christian Blanchette
MPI für Informatik –
Saarbrücken, DE
- Frédéric Blanqui
ENS – Cachan, FR
- Peter Brottveit Bock
IT University of
Copenhagen, DK
- Venanzio Capretta
University of Nottingham, GB
- Benjamin Delaware
Purdue University –
West Lafayette, US
- Gilles Dowek
INRIA & ENS Cachan, FR
- Catherine Dubois
ENSIIE – Evry, FR
- William M. Farmer
McMaster University –
Hamilton, CA
- Amy Felty
University of Ottawa, CA
- Thibault Gauthier
Universität Innsbruck, AT
- Frédéric Gilbert
ENS – Cachan, FR
- Georges Gonthier
INRIA Saclay –
Île-de-France, FR
- Stéphane Graham-Lengrand
Ecole Polytechnique –
Palaiseau, FR
- Hugo Herbelin
University Paris-Diderot, FR
- Olivier Hermant
Ecole des Mines de Paris, FR
- Guy Katz
Stanford University, US
- Chantal Keller
University of Paris Sud –
Orsay, FR
- Michael Kohlhase
Universität Erlangen-
Nürnberg, DE
- Ramana Kumar
Data61 / NICTA – Sydney, AU
- Dale Miller
INRIA Saclay –
Île-de-France, FR
- Alberto Momigliano
University of Milan, IT
- César A. Muñoz
NASA Langley – Hampton, US
- Adam Naumowicz
University of Białystok, PL
- Brigitte Pientka
McGill University –
Montreal, CA
- Elaine Pimentel
Federal University of
Rio Grande do Norte, BR
- Florian Rabe
Jacobs University Bremen, DE
- Claudio Sacerdoti Coen
University of Bologna, IT
- Ivan Scagnetto
University of Udine, IT
- Gert Smolka
Universität des Saarlandes, DE
- René Thiemann
Universität Innsbruck, AT
- Josef Urban
Czech Technical University –
Prague, CZ
- Laurent Voisin
SYSTEREL
Aix-en-Provence, FR
- Freek Wiedijk
Radboud University
Nijmegen, NL
- Bruno Woltzenlogel Paleo
Australian National University –
Canberra, AU



Computation over Compressed Structured Data

Edited by

Philip Bille¹, Markus Lohrey², Sebastian Maneth³, and Gonzalo Navarro⁴

1 Technical University of Denmark – Lyngby, DK, phbi@dtu.dk

2 Universität Siegen, DE, lohrey@eti.uni-siegen.de

3 University of Edinburgh, GB, smaneth@inf.ed.ac.uk

4 University of Chile – Santiago de Chile, CL, gnavarro@dcc.uchile.cl

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16431 “Computation over Compressed Structured Data”.

Seminar October 23–28, 2016 – <http://www.dagstuhl.de/16431>

1998 ACM Subject Classification Coding and Information Theory, Data Structures

Keywords and phrases algorithms on compressed structures, data compression, indexing, straight-line programs

Digital Object Identifier 10.4230/DagRep.6.10.99

Edited in cooperation with Fabian Peternek

1 Executive Summary

Philip Bille

Markus Lohrey

Sebastian Maneth

Gonzalo Navarro

License © Creative Commons BY 3.0 Unported license
© Philip Bille, Markus Lohrey, Sebastian Maneth, and Gonzalo Navarro

The Dagstuhl Seminar “Computation over Compressed Structured Data” took place from October 23rd to 28th, 2016. The aim was to bring together researchers from various research directions in data compression, indexing for compressed data, and algorithms for compressed data. Compression, and the ability to index and compute directly over compressed data, is a topic that is gaining importance as digitally stored data volumes are increasing at unprecedented speeds. In particular, the seminar focused on techniques for compressed *structured data*, i.e., string, trees, and graphs, where compression schemes can exploit complex structural properties to achieve strong compression ratios.

The seminar was meant to inspire the exchange of theoretical results and practical requirements related to compression of structured data, indexing, and algorithms for compressed structured data. The following specific points were addressed.

Encoding Data Structures. The goal is to encode data structures with the minimal number of bits needed to support only the desired operations, which is also called the effective entropy. The best known example of such an encoding is the $2n$ -bit structure that answers range minimum queries on a permutation of $[1, n]$, whose ordinary entropy is $n \log(n)$ bits. Determining the effective entropy and designing encodings that reach the effective



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

Computation over Compressed Structured Data, *Dagstuhl Reports*, Vol. 6, Issue 10, pp. 99–119

Editors: Philip Bille, Markus Lohrey, Sebastian Maneth, and Gonzalo Navarro



Dagstuhl Reports

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

entropy leads to challenging research problems in enumerative combinatorics, information theory, and data structures.

Computation-Friendly Compression. Existing state-of-the-art compression schemes encode data by extensive and convoluted references between pieces of information. This leads to strong compression guarantees, but often makes it difficult to efficiently perform compressed computation. Recent developments have moved towards designing more computation-friendly compression schemes that achieve both strong compression and allow for efficient computation. Precise bounds on the worst-case compression of these schemes are mostly missing so far.

Repetitive Text Collections. Many of the largest sequence collections that are arising are formed by many documents that are very similar to each other. Typical examples arise from version control systems, collaborative editing systems (wiki), or sequencing of genomes from the same species. Statistical-compression does not exploit this redundancy. Recently, compressed indexes based on grammar-based compressors have been developed for repetitive text collections. They achieve a considerable compression, but on the downside operations are much slower.

Recompression. Recompression is a new technique that was successfully applied for the approximation of smallest string grammars and to solve several algorithmic problems on grammar-compressed strings. Recently, recompression has been extended from strings to trees. The long list of problems that were solved in a relatively short period using recompression indicates that there exist more applications of recompression.

Graph Compression. A lot of recent work deals with succinct data structures for graphs and with graph compression, in particular for web and network graphs. At the same time, simple queries such as in- and out-neighbors can be executed efficiently on these structures. There is a wide range of important open problems and future work. For instance, there is a strong need to support more complex graph queries, like for instance regular path queries, on compressed graphs.

The seminar fully satisfied our expectations. The 41 participants from 16 countries (Algiers, Canada, Chile, Denmark, Finland, France, Germany, Great Britain, Ireland, Italy, Israel, Japan, Korea, Poland, Spain, and US) had been invited by the organizers to give survey talks about their recent research related to the topic of the seminar. The talks covered topics related to compression (e.g., grammar-based compression of string, trees, and graphs, Lempel-Ziv compression), indexing of compressed data (e.g., set-intersection, longest common extensions, labeling schemes), algorithms on compressed data (e.g., streaming, regular expression matching, parameterized matching) and covered a wide range of applications including databases, WWW, and bioinformatics. Most talks were followed by lively discussions. Smaller groups formed naturally which continued these discussions later.

We thank Schloss Dagstuhl for the professional and inspiring atmosphere. Such an intense research seminar is possible because Dagstuhl so perfectly meets all researchers' needs. For instance, elaborate research discussions in the evening were followed by local wine tasting or by heated sauna sessions.

2 Table of Contents

Executive Summary

Philip Bille, Markus Lohrey, Sebastian Maneth, and Gonzalo Navarro 99

Overview of Talks

| | |
|---|-----|
| Composite repetition-aware text indexing <i>Djamal Belazzougui</i> | 103 |
| Edit Distance: Sketching, Streaming and Document Exchange <i>Djamal Belazzougui</i> | 103 |
| Finger Search in Grammar-Compressed Strings <i>Philip Bille</i> | 104 |
| Towards Graph Re-compression <i>Stefan Böttcher</i> | 104 |
| Dynamic Relative Compression, Dynamic Partial Sums, and Substring Concatenation <i>Patrick Hagge Cording</i> | 105 |
| Compressed Affix Tree Representations <i>Rodrigo Cánovas</i> | 105 |
| Computing and Approximating the Lempel-Ziv-77 Factorization in Small Space <i>Johannes Fischer</i> | 106 |
| GLOUDS: Representing tree-like graphs <i>Johannes Fischer</i> | 106 |
| Queries on LZ-Bounded Encodings <i>Travis Gagie</i> | 107 |
| Distance and NCA labeling schemes for trees <i>Pawel Gawrychowski</i> | 107 |
| Querying regular languages over sliding-windows <i>Danny Hucke</i> | 108 |
| The smallest grammar problem revisited <i>Danny Hucke</i> | 109 |
| A Space-Optimal Grammar Compression <i>Tomohiro I</i> | 110 |
| Recompression <i>Artur Jez</i> | 110 |
| Linear Time String Indexing and Analysis in Small Space <i>Juha Kärkkäinen</i> | 111 |
| Dynamic Rank and Select Structures on Compressed Sequences <i>Yakov Nekrich</i> | 111 |
| Efficient Set Intersection Counting Algorithm for Text Similarity Measures <i>Patrick K. Nicholson</i> | 112 |
| Grammar-based Graph Compression <i>Fabian Peternek</i> | 112 |

| | |
|---|-----|
| In-place longest common extensions <i>Nicola Prezza</i> | 113 |
| Indexing in repetition-aware space <i>Nicola Prezza</i> | 113 |
| Encoding Data Structures <i>Rajeev Raman</i> | 114 |
| A Linear Time Algorithm for Seeds Computation <i>Wojciech Rytter</i> | 114 |
| Space-efficient graph algorithms <i>Srinivasa Rao Satti</i> | 115 |
| On the Complexity of Grammar-Based Compression over Fixed Alphabets <i>Markus Schmid</i> | 115 |
| Compressed parameterized pattern matching <i>Rahul Shah</i> | 115 |
| Streaming Pattern Matching <i>Tatiana Starikovskaya</i> | 116 |
| Quickscore: a fast algorithm to rank documents with additive ensembles of regression trees <i>Rossano Venturini</i> | 117 |
| Working groups | |
| LZ78 Construction in Little Main Memory Space <i>Diego Arroyuelo, Rodrigo Cánovas, Gonzalo Navarro, and Rajeev Raman</i> | 117 |
| Smaller Structures for Top- <i>k</i> Document Retrieval <i>Simon Gog, Julian Labeit, and Gonzalo Navarro</i> | 118 |
| More Efficient Representation of Web and Social Graphs by Combining GLOUDS with DSM <i>Cecilia Hernández Rivas, Johannes Fischer, Gonzalo Navarro, and Daniel Peters</i> . | 118 |
| Participants | 119 |

3 Overview of Talks

3.1 Composite repetition-aware text indexing

Djamal Belazzougui (CERIST – Algiers, DZ)

License © Creative Commons BY 3.0 Unported license
© Djamal Belazzougui

Joint work of Djamal Belazzougui, Fabio Cunial, Travis Gagie, Nicola Prezza, Mathieu Raffinot

Main reference D. Belazzougui, F. Cunial, T. Gagie, N. Prezza, M. Raffinot, “Composite Repetition-Aware Data Structures”, in Proc. of the Annual Symposium on Combinatorial Pattern Matching (CPM 2015), LNCS, Vol. 9133, pp. 26–39, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-19929-0_3

In highly repetitive strings, like collections of genomes from the same species, distinct measures of repetition all grow sublinearly in the length of the text, and indexes targeted to such strings typically depend only on one of these measures. We describe two data structures whose size depends on multiple measures of repetition at once, and that provide competitive tradeoffs between the time for counting and reporting all the exact occurrences of a pattern, and the space taken by the structure. The key component of our constructions is the run-length encoded BWT (RLBWT), which takes space proportional to the number of BWT runs: rather than augmenting RLBWT with suffix array samples, we combine it with data structures from LZ77 indexes, which take space proportional to the number of LZ77 factors, and with the compact directed acyclic word graph (CDAWG), which takes space proportional to the number of extensions of maximal repeats. The combination of CDAWG and RLBWT enables also a new representation of the suffix tree, whose size depends again on the number of extensions of maximal repeats, and that is powerful enough to support matching statistics and constant-space traversal.

3.2 Edit Distance: Sketching, Streaming and Document Exchange

Djamal Belazzougui (CERIST – Algiers, DZ)

License © Creative Commons BY 3.0 Unported license
© Djamal Belazzougui

Joint work of Djamal Belazzougui, Qin Zhang

Main reference D. Belazzougui, Q. Zhang, “Edit Distance: Sketching, Streaming and Document Exchange”, in Proc. of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2016), pp. 51–60, IEEE, 2016.

URL <http://dx.doi.org/10.1109/FOCS.2016.15>

We show that in the document exchange problem, where Alice holds a binary string x of length n and Bob holds another binary string y of length n , Alice can send Bob a message of size $O(K(\log^2 K + \log n))$ bits such that Bob can recover x using the message and his input y if the edit distance between x and y is no more than K , and output “error” otherwise. Both the encoding and decoding can be done in time $O(n + \text{poly}(K))$. This result significantly improves the previous communication bounds under polynomial encoding/decoding time. We also show that in the referee model, where Alice and Bob hold x and y respectively, they can compute sketches of x and y of sizes $\text{poly}(K \log n)$ bits (the encoding), and send to the referee, who can then compute the edit distance between x and y together with all the edit operations if the edit distance is no more than K , and output “error” otherwise (the decoding). To the best of our knowledge, this is the *first* result for sketching edit distance using $\text{poly}(K \log n)$ bits. Moreover, the encoding phase of our sketching algorithm can be performed by scanning the input string in one pass. Thus our sketching algorithm also

implies the “first” streaming algorithm for computing edit distance and all the edits exactly using $\text{poly}(K \log n)$ bits of space.

3.3 Finger Search in Grammar-Compressed Strings

Philip Bille (Technical University of Denmark – Lyngby, DK)

License  Creative Commons BY 3.0 Unported license
© Philip Bille

Joint work of Philip Bille, Anders Roy Christiansen, Patrick Hage Cording, Inge Li Gørtz

Main reference P. Bille, A. R. Christiansen, P. H. Cording, I. L. Gørtz, “Finger Search in Grammar-Compressed Strings”, in Proc. of the 36th Conf. on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2016), LIPIcs, Vol. 65, pp. 36:1–36:16, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.

URL <http://dx.doi.org/10.4230/LIPIcs.FSTTCS.2016.36>

Grammar-based compression, where one replaces a long string by a small context-free grammar that generates the string, is a simple and powerful paradigm that captures many popular compression schemes. Given a grammar, the random access problem is to compactly represent the grammar while supporting random access, that is, given a position in the original uncompressed string report the character at that position. In this paper we study the random access problem with the finger search property, that is, the time for a random access query should depend on the distance between a specified index f , called the *finger*, and the query index i . We consider both a static variant, where we first place a finger and subsequently access indices near the finger efficiently, and a dynamic variant where also moving the finger such that the time depends on the distance moved is supported. Let n be the size the grammar, and let N be the size of the string. For the static variant we give a linear space representation that supports placing the finger in $O(\log N)$ time and subsequently accessing in $O(\log D)$ time, where D is the distance between the finger and the accessed index. For the dynamic variant we give a linear space representation that supports placing the finger in $O(\log N)$ time and accessing and moving the finger in $O(\log D + \log \log N)$ time. Compared to the best linear space solution to random access, we improve a $O(\log N)$ query bound to $O(\log D)$ for the static variant and to $O(\log D + \log \log N)$ for the dynamic variant, while maintaining linear space. As an application of our results we obtain an improved solution to the longest common extension problem in grammar compressed strings. To obtain our results, we introduce several new techniques of independent interest, including a novel van Emde Boas style decomposition of grammars.

3.4 Towards Graph Re-compression

Stefan Böttcher (Universität Paderborn, DE)

License  Creative Commons BY 3.0 Unported license
© Stefan Böttcher

Re-compression of a compressed graph has the goal to find a stronger compression of the graph without full decompression of the graph. Having a tool for graph re-compression allows us to compress sub-graphs of a larger graph in parallel and to integrate several compressed sub-graphs afterwards by using the re-compression tool. While re-compression has been investigated for straight-line (SL) string-grammars and for SL binary tree-grammars, we present new ideas for the re-compression of SL tree-grammars with commutative terminal

nodes, i.e. SL tree-grammars representing partially unordered trees, and we extend these ideas to the re-compression of SL graph grammars representing graphs with labeled nodes and labeled edges. All our re-compression algorithms repetitively search a most frequent digram D and isolate and replace each digram occurrence of D by a new nonterminal N_D which is thereafter treated as a terminal. However, the re-compression approaches for strings, for ordered binary trees, for partially unordered trees, and for graphs differ in the following: what they consider a digram and how they define the key re-compression steps, i.e. counting (non-overlapping) digram occurrences and isolating digram occurrences.

3.5 Dynamic Relative Compression, Dynamic Partial Sums, and Substring Concatenation

Patrick Hagge Cording (Technical University of Denmark – Lyngby, DK)

License © Creative Commons BY 3.0 Unported license
© Patrick Hagge Cording

Joint work of Philip Bille, Patrick Hagge Cording, Inge Li Gørtz, Frederik Rye Skjoldjensen, Hjalte Wedel Vildhøj, Søren Vind

Main reference P. Bille, P. H. Cording, I. L. Gørtz, F. R. Skjoldjensen, H. W. Vildhøj, S. Vind, “Dynamic Relative Compression, Dynamic Partial Sums, and Substring Concatenation”, in Proc. of the 27th Int’l Symposium on Algorithms and Computation (ISAAC 2016), LIPIcs, Vol. 64, pp. 18:1–18:13, Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.

URL <http://dx.doi.org/10.4230/LIPIcs.ISAAC.2016.18>

Given a static reference string R and a source string S , a relative compression of S with respect to R is an encoding of S as a sequence of references to substrings of R . Relative compression schemes are a classic model of compression and have recently proved very successful for compressing highly-repetitive massive data sets such as genomes and web-data. We initiate the study of relative compression in a dynamic setting where the compressed source string S is subject to edit operations. The goal is to maintain the compressed representation compactly, while supporting edits and allowing efficient random access to the (uncompressed) source string. We present new data structures that achieve optimal time for updates and queries while using space linear in the size of the optimal relative compression, for nearly all combinations of parameters. We also present solutions for restricted and extended sets of updates. To achieve these results, we revisit the dynamic partial sums problem and the substring concatenation problem. We present new optimal or near optimal bounds for these problems. Plugging in our new results we also immediately obtain new bounds for the string indexing for patterns with wildcards problem and the dynamic text and static pattern matching problem.

3.6 Compressed Affix Tree Representations

Rodrigo Cánovas (University of Montpellier 2, FR)

License © Creative Commons BY 3.0 Unported license
© Rodrigo Cánovas

Joint work of Rodrigo Canovas, Eric Rivals

The Suffix Tree, a crucial and versatile data structure for string analysis of large texts, is often used in pattern matching and in bioinformatics applications. The Affix Tree generalizes the Suffix Tree in that it supports full tree functionalities in both search directions. The

bottleneck of Affix Trees is their space requirement for storing the data structure. Here, we discuss existing representations and classify them into two categories: Synchronous and Asynchronous. We design Compressed Affix Tree indexes in both categories and explored how to support all tree operations bidirectionally. This work compares alternative approaches for compressing the Affix Tree, measuring their space and time trade-offs for different operations. Moreover, to our knowledge, this is the first work that compares all Compressed Affix Tree implementations offering a practical benchmark for this structure.

3.7 Computing and Approximating the Lempel-Ziv-77 Factorization in Small Space

Johannes Fischer (TU Dortmund, DE)

License  Creative Commons BY 3.0 Unported license
© Johannes Fischer

The Lempel-Ziv-77 algorithm greedily factorizes a text of length n into z maximal substrings that have previous occurrences, which is particularly useful for text compression. We review two recent algorithms for this task:

1. A linear-time algorithm using essentially only one integer array of length n in addition to the text. (Joint work with Tomohiro I and Dominik Köppl.)
2. An even more space-conscious algorithm using $O(z)$ space, computing a 2-approximation of the LZ77 parse in $O(n \lg n)$ time w.h.p. (Joint work with Travis Gagie, Pawel Gawrychowski and Tomasz Kociumaka.)

3.8 GLOUDS: Representing tree-like graphs

Johannes Fischer (TU Dortmund, DE)

License  Creative Commons BY 3.0 Unported license
© Johannes Fischer

Joint work of Johannes Fischer, Daniel Peters

Main reference J. Fischer, D. Peters, “GLOUDS: Representing tree-like graphs”, *Journal of Discrete Algorithms*, Vol. 36, pp. 39–49, Elsevier, 2016.

URL <http://dx.doi.org/10.1016/j.jda.2015.10.004>

The Graph Level Order Unary Degree Sequence (GLOUDS) is a new succinct data structure for directed graphs that are “tree-like,” in the sense that the number of “additional” edges (w.r.t. a spanning tree) is not too high. The algorithmic idea is to represent a BFS-spanning tree of the graph (consisting of n nodes) with a well known succinct data structure for trees, named LOUDS, and enhance it with additional information that accounts for the non-tree edges. In practical tests, our data structure performs well for graphs containing up to $m = 5n$ edges, while still having competitive running times for listing adjacent nodes.

3.9 Queries on LZ-Bounded Encodings

Travis Gagie (Universidad Diego Portales, CL)

License © Creative Commons BY 3.0 Unported license
© Travis Gagie

Joint work of Djamel Belazzougui, Travis Gagie, Pawel Gawrychowski, Juha Kärkkäinen, Alberto Ordóñez, Simon J. Puglisi, Yasuo Tabei

Main reference D. Belazzougui, T. Gagie, P. Gawrychowski, J. Kärkkäinen, A. Ordóñez, S. J. Puglisi, Y. Tabei, “Queries on LZ-Bounded Encodings”, Data Compression Conference (DCC 2015), pp. 83–92, IEEE, 2015.

URL <http://dx.doi.org/10.1109/DCC.2015.69>

We describe a data structure that stores a strings in space similar to that of its Lempel-Ziv encoding and efficiently supports access, rank and select queries. These queries are fundamental for implementing succinct and compressed data structures, such as compressed trees and graphs. We show that our data structure can be built in a scalable manner and is both small and fast in practice compared to other data structures supporting such queries.

3.10 Distance and NCA labeling schemes for trees

Pawel Gawrychowski (University of Wroclaw, PL)

License © Creative Commons BY 3.0 Unported license
© Pawel Gawrychowski

Joint work of Ofer Freedman, Pawel Gawrychowski, Jakub Łopuszański, Patrick K. Nicholson, Oren Weimann

Main reference O. Freedman, P. Gawrychowski, P. K. Nicholson, O. Weimann, “Optimal Distance Labeling Schemes for Trees”, arXiv:1608.00212v1 [cs.DS], 2016.

URL <https://arxiv.org/abs/1608.00212v1>

Labeling schemes seek to assign a short label to each vertex in a graph, so that a function on two nodes (such as distance or adjacency) can be computed by examining their labels alone. This is particularly desirable in distributed settings, where nodes are often processed using only some locally stored data. Recently, with the rise in popularity of distributed computing platforms such as Spark and Hadoop, labeling schemes have found renewed interest. For the particular case of trees, the most natural functions are distance, ancestry, adjacency, and nearest common ancestor. We design improved labeling schemes for distance and nearest common ancestor.

For arbitrary distances, we show how to assign labels of $1/4 \log^2 n + o(\log^2 n)$ bits to every node so that we can determine the distance between any two nodes given only their two labels. This closes the line of research initiated by Gavaille et al. [SODA '01] who gave an $O(\log^2 n)$ bits upper bound and a $1/8 \log^2 n - O(\log n)$ bits lower bound, and followed by Alstrup et al. [ICALP '16] who gave a $1/2 \log^2 n + O(\log n)$ bits upper bound and a $1/4 \log^2 n - O(\log n)$ bits lower bound.

Next, for distances bounded by k , we show how to construct labels whose length is the minimum between $\log n + O(k \log(\log n/k))$ and $O(\log n \cdot \log(k/\log n))$. The query time in both cases is constant. We complement our upper bounds with almost tight lower bounds of $\log n + \Omega(k \log(\log n/(k \log k)))$ and $\Omega(\log n \cdot \log(k/\log n))$. Finally, we consider $(1 + \varepsilon)$ -approximate distances. We prove an $O(\log(1/\varepsilon) \cdot \log n)$ upper bound and a matching $\Omega(\log(1/\varepsilon) \cdot \log n)$ lower bound. This improves the recent $O(1/\varepsilon \cdot \log n)$ upper bound of Alstrup et al. [ICALP '16].

For nearest common ancestor, the known upper bounds are in the $O(\log n)$ regime. We significantly improve the constant factor, and in particular go below 2 for the case of binary trees.

3.11 Querying regular languages over sliding-windows

Danny Hucke (Universität Siegen, DE)

License  Creative Commons BY 3.0 Unported license
© Danny Hucke

Joint work of Moses Ganardi, Danny Hucke, Markus Lohrey

We study the space complexity of querying regular languages over data streams in the sliding window model. The algorithm has to answer at any point of time whether the content of the sliding window belongs to a fixed regular language. A trichotomy is shown: For every regular language the optimal space requirement is either in $\Theta(n)$, $\Theta(\log n)$, or constant, where n is the size of the sliding window.

References

- 1 Charu C. Aggarwal. *Data Streams – Models and Algorithms*. Springer, 2007.
- 2 Arvind Arasu and Gurmeet Singh Manku. Approximate counts and quantiles over sliding windows. In *Proceedings of PODS 2004*, pages 286–296. ACM, 2004.
- 3 Brian Babcock, Mayur Datar, Rajeev Motwani, and Liadan O’Callaghan. Maintaining variance and k-medians over data stream windows. In *Proceedings of PODS 2003*, pages 234–243. ACM, 2003.
- 4 Ajesh Babu, Nutan Limaye, Jaikumar Radhakrishnan, and Girish Varma. Streaming algorithms for language recognition problems. *Theor. Comput. Sci.*, 494:13–23, 2013.
- 5 Vladimir Braverman. Sliding window algorithms. In *Encyclopedia of Algorithms*, pages 2006–2011. 2016.
- 6 Vladimir Braverman, Rafail Ostrovsky, and Carlo Zaniolo. Optimal sampling from sliding windows. *J. Comput. Syst. Sci.*, 78(1):260–272, 2012.
- 7 Michael S. Crouch, Andrew McGregor, and Daniel Stubbs. Dynamic graphs in the sliding-window model. In *Proceedings of ESA 2013*, volume 8125 of *Lecture Notes in Computer Science*, pages 337–348. Springer, 2013.
- 8 Mayur Datar, Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Maintaining stream statistics over sliding windows. *SIAM J. Comput.*, 31(6):1794–1813, 2002.
- 9 Manfred Droste, Werner Kuich, and Heiko Vogler. *Handbook of Weighted Automata*. Springer, 2009.
- 10 Gudmund Skovbjerg Frandsen, Peter Bro Miltersen, and Sven Skyum. Dynamic word problems. *J. ACM*, 44(2):257–271, 1997.
- 11 Lukasz Golab and M. Tamer Özsu. Processing sliding window multi-joins in continuous queries over data streams. In *Proceedings of VLDB 2003*, pages 500–511. Morgan Kaufmann, 2003.
- 12 Markus Holzer and Barbara König. On deterministic finite automata and syntactic monoid size. *Theor. Comput. Sci.*, 327(3):319–347, 2004.
- 13 Howard Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhäuser, Boston, Basel, Berlin, 1994.

3.12 The smallest grammar problem revisited

Danny Hucke (*Universität Siegen, DE*)

License  Creative Commons BY 3.0 Unported license
© Danny Hucke

Joint work of Danny Hucke, Markus Lohrey, Carl Philipp Reh

In a seminal paper of Charikar et al. on the smallest grammar problem, the authors derive upper and lower bounds on the approximation ratios for several grammar-based compressors, but in all cases there is a gap between the lower and upper bound. Here we close the gaps for LZ78 and BISECTION by showing that the approximation ratio of LZ78 is $\Theta((n/\log n)^{2/3})$, whereas the approximation ratio of BISECTION is $\Theta((n/\log n)^{1/2})$.

References

- 1 J. Arpe and R. Reischuk. On the complexity of optimal grammar-based compression. In *Proc. DCC 2006*, pages 173–182. IEEE Computer Society, 2006.
- 2 J. Berstel and S. Brlek. On the length of word chains. *Inf. Process. Lett.*, 26(1):23–28, 1987.
- 3 K. Casel, H. Fernau, S. Gaspers, B. Gras, and M. L. Schmid. On the complexity of grammar-based compression over fixed alphabets. In *Proc. ICALP 2016*, Lecture Notes in Computer Science. Springer, 1996. to appear.
- 4 M. Charikar, E. Lehman, A. Lehman, D. Liu, R. Panigrahy, M. Prabhakaran, A. Sahai, and A. Shelat. The smallest grammar problem. *IEEE Trans. Inf. Theory*, 51(7):2554–2576, 2005.
- 5 A. A. Diwan. A new combinatorial complexity measure for languages. Tata Institute, Bombay, India, 1986.
- 6 L. Gasieniec, M. Karpinski, W. Plandowski, and W. Rytter. Efficient algorithms for Lempel-Ziv encoding (extended abstract). In *Proc. SWAT 1996*, volume 1097 of *Lecture Notes in Computer Science*, pages 392–403. Springer, 1996.
- 7 A. Jež. Approximation of grammar-based compression via recompression. In *Proc. CPM 2013*, volume 7922 of *LNCS*, pages 165–176. Springer, 2013.
- 8 J. C. Kieffer and E.-H. Yang. Grammar-based codes: A new class of universal lossless source codes. *IEEE Trans. Inf. Theory*, 46(3):737–754, 2000.
- 9 J. C. Kieffer, E.-H. Yang, G. J. Nelson, and P. C. Cosman. Universal lossless compression via multilevel pattern matching. *IEEE Trans. Inf. Theory*, 46(4):1227–1245, 2000.
- 10 N. J. Larsson and A. Moffat. Offline dictionary-based compression. In *Proc. DCC 1999*, pages 296–305. IEEE Computer Society, 1999.
- 11 M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications, Third Edition*. Springer, 2008.
- 12 M. Lohrey. *The Compressed Word Problem for Groups*. Springer, 2014.
- 13 C. G. Nevill-Manning and I. H. Witten. Identifying hierarchical structure in sequences: A linear-time algorithm. *J. Artif. Intell. Res.*, 7:67–82, 1997.
- 14 W. Rytter. Application of Lempel-Ziv factorization to the approximation of grammar-based compression. *Theor. Comput. Sci.*, 302(1–3):211–222, 2003.
- 15 J. A. Storer and T. G. Szymanski. Data compression via textual substitution. *J. ACM*, 29(4):928–951, 1982.
- 16 Y. Tabei, Y. Takabatake, and H. Sakamoto. A succinct grammar compression. In *Proc. CPM 2013*, volume 7922 of *Lecture Notes in Computer Science*, pages 235–246. Springer, 2013.
- 17 J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inf. Theory*, 24(5):530–536, 1977.

3.13 A Space-Optimal Grammar Compression

Tomohiro I (*Kyushu Institute of Technology, JP*)

License  Creative Commons BY 3.0 Unported license
© Tomohiro I

Joint work of Hiroshi Sakamoto, Tomohiro I, Yoshimasa Takabatake

A grammar compression is a context-free grammar (CFG) deriving a single string deterministically. For a CFG G in Chomsky normal form of n -symbol, it is known that an information-theoretic lower bound to represent G is $\lg n! + n + o(n)$ bits. Although a fully-online algorithm for constructing a succinct G of at most $n \lg(n + \sigma) + o(n \lg(n + \sigma))$ bits was proposed for the number σ of alphabets and the number n of variables, the optimization of the working space has remained open. We achieve the smallest $n \lg(n + \sigma) + o(n \lg(n + \sigma))$ bits of working space while preserving $O(N \frac{\lg n}{\lg \lg n})$ amortized compression time for the length N of the input string received so far.

3.14 Recompression

Artur Jez (*University of Wrocław, PL*)

License  Creative Commons BY 3.0 Unported license
© Artur Jez

In this talk I will survey the recompression technique. It is based on applying simple compression operations (replacement of pairs of two different letters by a new letter and replacement of maximal repetition of a letter by a new symbol) applied to strings. The strings in question are given in a compressed way: each is represented as a context free grammar generating exactly one string, called SLP in the following. The operations are conceptually applied on the strings but they are actually performed directly on the compressed representation. For instance, when we want to replace ab in the string and the grammar have a production $X \rightarrow aY$ and the string generated by Y is bw , then we alter the rule of Y so that it generates w and replace Y with bY in all rules. In this way the rule is $X \rightarrow abY$ and so ab can be replaced. In this way we are interested mostly in the way the string is compressed rather than the string and its combinatorial properties.

The proposed method turned out to be surprisingly efficient and applicable in various scenarios: it can be used to test the equality of SLPs in time $O(n \log N)$, where n is the size of the SLP and N the length of the generated string. It can be also used to approximate the smallest SLP for a given string, with the approximation ratio $O(\log(n/g))$, where n is the length of the string and g the size of the smallest SLP for this string. Furthermore, it works also when the strings are given by more implicit representations: as solutions to word equations. This approach can be also generalized to trees and most of the results extend from the string to tree setting.

3.15 Linear Time String Indexing and Analysis in Small Space

Juha Kärkkäinen (University of Helsinki, FI)

License © Creative Commons BY 3.0 Unported license
© Juha Kärkkäinen

The field of succinct data structures has flourished over the last 16 years. Starting from the compressed suffix array (CSA) by Grossi and Vitter (STOC 2000) and the FM-index by Ferragina and Manzini (FOCS 2000), a number of generalizations and applications of string indexes based on the Burrows-Wheeler transform (BWT) have been developed, all taking an amount of space that is close to the input size in bits. In many large-scale applications, the construction of the index and its usage need to be considered as one unit of computation. Efficient string indexing and analysis in small space lies also at the core of a number of primitives in the data-intensive field of high-throughput DNA sequencing. We report the following advances in string indexing and analysis. We show that the BWT of a string $T \in \{1, \dots, \sigma\}^n$ can be built in deterministic $O(n)$ time using just $O(n \log \sigma)$ bits of space, where $\sigma \leq n$. Within the same time and space budget, we can build an index based on the BWT that allows one to enumerate all the internal nodes of the suffix tree of T . Many fundamental string analysis problems can be mapped to such enumeration, and can thus be solved in deterministic $O(n)$ time and in $O(n \log \sigma)$ bits of space from the input string. We also show how to build many of the existing indexes based on the BWT, such as the CSA, the compressed suffix tree (CST), and the bidirectional BWT index, in randomized $O(n)$ time and in $O(n \log \sigma)$ bits of space. The previously fastest construction algorithms for BWT, CSA and CST, which used $O(n \log \sigma)$ bits of space, took $O(n \log \log \sigma)$ time for the first two structures, and $O(n \log^\varepsilon n)$ time for the third, where ε is any positive constant. Contrary to the state of the art, our bidirectional BWT index supports every operation in constant time per element in its output.

3.16 Dynamic Rank and Select Structures on Compressed Sequences

Yakov Nekrich (University of Waterloo, CA)

License © Creative Commons BY 3.0 Unported license
© Yakov Nekrich

Joint work of Gonzalo Navarro, Yakov Nekrich, Ian Munro

We consider the problem of storing a fully-dynamic string S in compressed form. Our representation supports insertions and deletions of symbols and answers three fundamental queries: $\text{access}(i, S)$ returns the i -th symbol in S , $\text{rank}_a(i, S)$ counts how many times a symbol a occurs among the first i positions in S , and $\text{select}_a(i, S)$ finds the position where a symbol a occurs for the i -th time. Data structures supporting rank, select, and access queries are used in many compressed data structures and algorithms that work on string data.

We give an overview of previous results and describe two state-of-the-art solutions for this problem.

References

- 1 J. Ian Munro, Yakov Nekrich. *Compressed Data Structures for Dynamic Sequences*. ESA 2015:891–902
- 2 Gonzalo Navarro, Yakov Nekrich. *Optimal Dynamic Sequence Representations*. SIAM J. Comput. 43(5):1781–1806 (2014)

3.17 Efficient Set Intersection Counting Algorithm for Text Similarity Measures

Patrick K. Nicholson (Bell Labs – Dublin, IE)

License © Creative Commons BY 3.0 Unported license
© Patrick K. Nicholson

Joint work of Preethi Lahoti, Patrick K. Nicholson, Bilyana Taneva

Main reference P. Lahoti, P. K. Nicholson, B. Taneva, “Efficient Set Intersection Counting Algorithm for Text Similarity Measures”, in Proc. of the 19th Workshop on Algorithm Engineering & Experiments (ALENEX 2017), pp. 146–158, SIAM, 2017.

URL <http://dx.doi.org/10.1137/1.9781611974768.12>

Set intersection counting appears as a subroutine in many techniques used in natural language processing, in which similarity is often measured as a function of document cooccurrence counts between pairs of noun phrases or entities. Such techniques include clustering of text phrases and named entities, topic labeling, entity disambiguation, sentiment analysis, and search for synonyms.

These techniques can have real-time constraints that require very fast computation of thousands of set intersection counting queries with little space overhead and minimal error. On one hand, while sketching techniques for approximate intersection counting exist and have very fast query time, many have issues with accuracy, especially for pairs of lists that have low Jaccard similarity. On the other hand, space-efficient computation of exact intersection sizes is particularly challenging in real-time.

In this paper, we show how an efficient space-time trade-off can be achieved for exact set intersection counting, by combining state-of-the-art algorithms with precomputation and judicious use of compression. In addition, we show that the performance can be further improved by combining the best aspects of these algorithms. We present experimental evidence that real-time computation of exact intersection sizes is feasible with low memory overhead: we improve the mean query time of baseline approaches by over a factor of 100 using a data structure that takes merely twice the size of an inverted index. Overall, in our experiments, we achieve running times within the same order of magnitude as well-known approximation techniques.

3.18 Grammar-based Graph Compression

Fabian Peternek (University of Edinburgh, GB)

License © Creative Commons BY 3.0 Unported license
© Fabian Peternek

Joint work of Sebastian Maneth, Fabian Peternek

Main reference S. Maneth, F. Peternek, “Compressing Graphs by Grammars”, in Proc. of the 32nd Int’l Conf. on Data Engineering (ICDE 2016), IEEE, 2016.

URL <http://dx.doi.org/10.1109/ICDE.2016.7498233>

This talk considers a method for compressing graphs into a smaller graph grammar based on the RePair compression scheme. We start by defining the necessary notion of context-free hyperedge replacement grammars. We then extend RePair to graphs using this grammar formalism and discuss how the problem of finding non-overlapping occurrences appears more difficult on graphs than on strings and trees.

We also give some intuition on graphs that are difficult to compress with HR grammars and present some experimental results based on a proof-of-concept implementation.

Finally a short overview on two linear time speed-up algorithms for such compressed graph grammars is presented, namely reachability and regular path queries.

3.19 In-place longest common extensions

Nicola Prezza (*University of Udine, IT*)

License © Creative Commons BY 3.0 Unported license
© Nicola Prezza

Main reference N. Prezza, “In-Place Longest Common Extensions”, arXiv:1608.05100v8 [cs.DS], 2016.

URL <https://arxiv.org/abs/1608.05100v8>

A Longest Common Extension (LCE) query returns the length of the longest common prefix between any two text suffixes. In this talk I present a deterministic data structure having the exact same size of the text (i.e. without additional low-order terms in its space occupancy) and supporting LCE queries in logarithmic time and text extraction in optimal time. Importantly, the structure can be built in-place: we can replace the text with the structure while using only $O(1)$ memory words of additional space during construction. This result has interesting implications: I show the first sub-quadratic in-place algorithms to compute the LCP array and to solve the sparse suffix sorting problem, and a new in-place suffix array construction algorithm.

3.20 Indexing in repetition-aware space

Nicola Prezza (*University of Udine, IT*)

License © Creative Commons BY 3.0 Unported license
© Nicola Prezza

Joint work of Djamel Belazzougui, Fabio Cunial, Travis Gagie, Nicola Prezza, Mathieu Raffinot, Alberto Policriti
Main reference D. Belazzougui, F. Cunial, T. Gagie, N. Prezza, M. Raffinot, “Composite repetition-aware data structures”, in Proc. of the 26th Ann. Symp. on Combinatorial Pattern Matching (CPM 2015), LNCS, Vol. 9133, pp. 26–39, Springer, 2015; pre-print available as arXiv:1502.05937v2 [cs.DS].

URL http://dx.doi.org/10.1007/978-3-319-19929-0_3

URL <https://arxiv.org/abs/1502.05937v2>

Full-text indexes based on the Lempel-Ziv factorization (LZ77) and on the run-length compressed Burrows-Wheeler transform (RLBWT) achieve strong compression rates, but their construction in small (compressed) space is still an open problem. In this talk, I present an algorithm to compute LZ77 in space proportional to the number of equal-letter runs in the Burrows-Wheeler transform. This result implies an asymptotically optimal-space construction algorithm for the lz-rlbwt index: an index combining LZ77 with RLBWT able to achieve exponential compression while supporting sub-quadratic time count and locate operations. I moreover present DYNAMIC: a C++ library implementing dynamic compressed data structures. This library has been used in conjunction with SDSL to implement all the discussed algorithms and indexes. I will conclude discussing different practical implementations of the lz-rlbwt index and presenting experimental results comparing our LZ77 factorization algorithm and the lz-rlbwt index with the state of the art.

3.21 Encoding Data Structures

Rajeev Raman (University of Leicester, GB)

License © Creative Commons BY 3.0 Unported license
© Rajeev Raman

Joint work of This is a survey talk based on papers by many authors.

Main reference R. Raman, “Encoding Data Structures”, in Proc. of the 9th Int’l Workshop on Algorithms and Computation (WALCOM’15), LNCS, Vol. 8973, pp. 1–7, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-15612-5_1

Driven by the increasing need to analyze and search for complex patterns in very large data sets, the area of compressed and succinct data structures has grown rapidly in the last 10–15 years. Such data structures have very low memory requirements, allowing them to fit into the main memory of a computer, which in turn avoids expensive computation on hard disks.

This talk will focus on a topic that has become popular recently: encoding “the data structure” itself. Some data structuring problems involve supporting queries on data, but the queries that need to be supported do not allow the original data to be deduced from the queries. This presents opportunities to obtain space savings even when the data is incompressible, by extracting only the information needed to answer the queries when pre-processing the data, and then deleting the data. The minimum information needed to answer the queries is called the *effective entropy* of the problem: precisely determining the effective entropy leads to interesting combinatorial problems.

This survey talk is a slightly longer version of a talk given in Dagstuhl seminar 16101 (Data Structures and Advanced Models of Computation on Big Data) and was given at the request of the organizers.

3.22 A Linear Time Algorithm for Seeds Computation

Wojciech Rytter (University of Warsaw, PL)

License © Creative Commons BY 3.0 Unported license
© Wojciech Rytter

Joint work of Tomasz Kociumaka, Marcin Kubica, Jakub Radoszewski, Wojciech Rytter, Tomasz Waleń

Main reference T. Kociumaka, M. Kubica, J. Radoszewski, W. Rytter, T. Waleń, “A linear time algorithm for seeds computation”, in Proc. of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2012), pp. 1095–1112, ACM, 2012.

URL <http://dl.acm.org/citation.cfm?id=2095116.2095202>

A seed in a word is a relaxed version of a period. We show a linear-time algorithm computing a compact representation of linear size of all the seeds of a word (though the set of distinct seeds could be quadratic). In particular, the algorithm computes the shortest seed and the number of seeds. Our approach is based on combinatorial relations between seeds and a variant of the LZ-factorization (used here for the first time in context of seeds). In the previous papers the compact representation of seeds consisted of two independent representations based on the suffix tree of the word and the suffix tree of the reverse of the word. Our another contribution is a new compact representation of all seeds which avoids dealing with reversals of the word.

3.23 Space-efficient graph algorithms

Srinivasa Rao Satti (Seoul National University, KR)

License © Creative Commons BY 3.0 Unported license
© Srinivasa Rao Satti

Joint work of Sankardeep Chakraborty, Anish Mukherjee, Srinivasa Rao Satti, Venkatesh Raman

We reconsider various graph algorithms in the settings where there is only a limited amount of memory available, apart from the input representation. These settings are motivated by the need to run such algorithms on limited-memory devices, as well as to capture the essential features of some of the latest memory technologies. The talk presents a brief overview of some of the recent developments in this area, and gives pointers to some future directions.

3.24 On the Complexity of Grammar-Based Compression over Fixed Alphabets

Markus Schmid (Universität Trier, DE)

License © Creative Commons BY 3.0 Unported license
© Markus Schmid

Joint work of Katrin Casel, Henning Fernau, Serge Gaspers, Benjamin Gras, Markus L. Schmid

It is shown that the shortest-grammar problem remains NP-complete if the alphabet is fixed and has a size of at least 24 (which settles an open question). On the other hand, this problem can be solved in polynomial-time, if the number of nonterminals is bounded, which is shown by encoding the problem as a problem on graphs with interval structure. Furthermore, we present an $O(3^n)$ exact exponential-time algorithm, based on dynamic programming. Similar results are also given for 1-level grammars, i.e., grammars for which only the start rule contains nonterminals on the right side (thus, investigating the impact of the “hierarchical depth” on the complexity of the shortest-grammar problem).

3.25 Compressed parameterized pattern matching

Rahul Shah (Louisiana State University – Baton Rouge, US)

License © Creative Commons BY 3.0 Unported license
© Rahul Shah

Joint work of Arnab Ganguly, Rahul Shah, Sharma Thankachan

Main reference A. Ganguly, R. Shah, S. Thankachan, “Parameterized Pattern Matching – Succinctly”, arXiv:1603.07457v2 [cs.DS], 2016.

URL <https://arxiv.org/abs/1603.07457v2>

The fields of succinct data structures and compressed text indexing have seen quite a bit of progress over the last two decades. An important achievement, primarily using techniques based on the Burrows-Wheeler Transform (BWT), was obtaining the full functionality of the suffix tree in the optimal number of bits. A crucial property that allows the use of BWT for designing compressed indexes is *order-preserving suffix links*. Specifically, the relative order between two suffixes in the subtree of an internal node is same as that of the suffixes obtained by truncating the first character of the two suffixes. Unfortunately, in many variants of the text-indexing problem, for e.g., parameterized pattern matching, 2D pattern matching, and order-isomorphic pattern matching, this property does not hold.

Consequently, the compressed indexes based on BWT do not directly apply. Furthermore, a compressed index for any of these variants has been elusive throughout the advancement of the field of succinct data structures. We achieve a positive breakthrough on one such problem, namely the *Parameterized Pattern Matching* problem.

Let T be a text that contains n characters from an alphabet Σ , which is the union of two disjoint sets: Σ_s containing static characters (s-characters) and Σ_p containing parameterized characters (p-characters). A pattern P (also over Σ) matches an equal-length substring S of T iff the s-characters match exactly, and there exists a one-to-one function that renames the p-characters in S to that in P . The task is to find the starting positions (occurrences) of all such substrings S . Previous index [Baker, STOC 1993], known as *Parameterized Suffix Tree*, requires $\Theta(n \log n)$ bits of space, and can find all occ occurrences in time $O(|P| \log \sigma + occ)$, where $\sigma = |\Sigma|$. We introduce an $n \log \sigma + O(n)$ -bit index with $O(|P| \log \sigma + occ \cdot \log n \log \sigma)$ query time. At the core, lies a new BWT-like transform, which we call the *Parameterized Burrows-Wheeler Transform* (pBWT). The techniques are extended to obtain a succinct index for the *Parameterized Dictionary Matching* problem of Idury and Schäffer [CPM, 1994].

3.26 Streaming Pattern Matching

Tatiana Starikovskaya (University Paris-Diderot, FR)

License  Creative Commons BY 3.0 Unported license
© Tatiana Starikovskaya

Joint work of Raphaël Clifford, Allyx Fontaine, Ely Porat, Benjamin Sach, Tatiana Starikovskaya

In the streaming model of computation we assume that the data arrives sequentially, one data item at a time. The goal is to develop algorithms that process the data on the fly while using as little space as possible. We give a survey of recent results for the pattern matching problem in this model. We first review the main ideas behind the algorithm for exact pattern matching problem given by Porat and Porat and show how to extend them to the case of several patterns (dictionary matching). We then proceed to the approximate pattern matching problem under Hamming distance. In this problem we must compute the Hamming distance for all alignments of the given pattern and the text. We first consider the famous variant of this problem called the k -mismatch problem, where we are only interested in small Hamming distances (smaller than a given threshold k). Finally, we consider a problem of computing all Hamming distances and present an approximate algorithm for it.

References

- 1 R. Clifford, A. Fontaine, E. Porat, B. Sach, T. Starikovskaya. *The k -mismatch problem revisited*. SODA 2016:2039–2052. DOI: <http://dx.doi.org/10.1137/1.9781611974331.ch142>
- 2 R. Clifford, T. Starikovskaya. *Approximate Hamming Distance in a Stream*. ICALP 2016: 20:1–20:14 DOI: <http://dx.doi.org/10.4230/LIPIcs.ICALP.2016.20>
- 3 R. Clifford, A. Fontaine, E. Porat, B. Sach, T. Starikovskaya. *Dictionary Matching in a Stream*. ESA 2015:361–372 DOI: http://dx.doi.org/10.1007/978-3-662-48350-3_31

3.27 Quickscore: a fast algorithm to rank documents with additive ensembles of regression trees

Rossano Venturini (*University of Pisa, IT*)

License © Creative Commons BY 3.0 Unported license

© Rossano Venturini

Joint work of Claudio Lucchese, Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, Nicola Tonello, Rossano Venturini

Main reference C. Lucchese, F.M. Nardini, S. Orlando, R. Perego, N. Tonello, R. Venturini, “QuickScorer: a Fast Algorithm to Rank Documents with Additive Ensembles of Regression Trees”, in Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR’15), pp. 73–82, ACM, 2015.

URL <http://dx.doi.org/10.1145/2766462.2767733>

Learning-to-Rank models based on additive ensembles of regression trees have proven to be very effective for ranking query results returned by Web search engines, a scenario where quality and efficiency requirements are very demanding. Unfortunately, the computational cost of these ranking models is high. Thus, several works already proposed solutions aiming at improving the efficiency of the scoring process by dealing with features and peculiarities of modern CPUs and memory hierarchies. In this paper, we present QuickScorer, a new algorithm that adopts a novel bitvector representation of the tree-based ranking model, and performs an interleaved traversal of the ensemble by means of simple logical bitwise operations. The performance of the proposed algorithm are unprecedented, due to its cacheaware approach, both in terms of data layout and access patterns, and to a control flow that entails very low branch mis-prediction rates. The experiments on real Learning-to-Rank datasets show that QuickScorer is able to achieve speedups over the best state-of-the-art baseline ranging from 2x to 6.5x.

4 Working groups

4.1 LZ78 Construction in Little Main Memory Space

Diego Arroyuelo (*TU Federico Santa María – Valparaíso, CL*), Rodrigo Cánovas (*University of Montpellier 2, FR*), Gonzalo Navarro (*University of Chile – Santiago de Chile, CL*), and Rajeev Raman (*University of Leicester, GB*)

License © Creative Commons BY 3.0 Unported license

© Diego Arroyuelo, Rodrigo Cánovas, Gonzalo Navarro, and Rajeev Raman

Joint work of Diego Arroyuelo, Rodrigo Cánovas, Gonzalo Navarro, Andreas Poyias, Rajeev Raman

Main reference A. Poyias, R. Raman, “Improved Practical Compact Dynamic Tries”, in Proc. of the 22nd Int’l Symposium on String Processing and Information Retrieval (SPIRE 2015), LNCS, Vol. 9309, pp. 324–336, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-23826-5_31

We report on ongoing work aiming to do the LZ78 parsing of a text T $[1..n]$ over alphabet $[1..\sigma]$ in linear randomized time, using only $O(z \log \sigma)$ bits of main memory, while reading the input text from disk and writing the compressed text to disk. The text can also be decompressed within the same main memory usage.

4.2 Smaller Structures for Top- k Document Retrieval

Simon Gog (KIT – Karlsruher Institut für Technologie, DE), Julian Labeit, and Gonzalo Navarro (University of Chile – Santiago de Chile, CL)

License © Creative Commons BY 3.0 Unported license

© Simon Gog, Julian Labeit, and Gonzalo Navarro

Main reference J. Labeit, S. Gog, “Elias-Fano meets Single-Term Top- k Document Retrieval”, in Proc. of the 19th Workshop on Algorithm Engineering & Experiments (ALENEX 2017), pp. 135–145, SIAM, 2017.

URL <http://dx.doi.org/10.1137/1.9781611974768.11>

In a recent paper (main reference) Labeit and Gog improve upon the space of a previous fast top- k document retrieval index by Gog and Navarro [Improved Single-Term Top- k Document Retrieval. Proc. ALENEX’15, pages 24-32], by sharply reducing the information stored about document identifiers. We now plan to further reduce the space, without hopefully affect the time too much, by removing the information on frequencies, which is the largest remaining component of the index. We plan to replace this information with a small index per document, using a previously developed (and unpublished) technique by Navarro to store many small document indexes within little space overhead.

4.3 More Efficient Representation of Web and Social Graphs by Combining GLOUDS with DSM

Cecilia Hernández Rivas (University of Concepción, CL), Johannes Fischer (TU Dortmund, DE), Gonzalo Navarro (University of Chile – Santiago de Chile, CL), and Daniel Peters

License © Creative Commons BY 3.0 Unported license

© Cecilia Hernández Rivas, Johannes Fischer, Gonzalo Navarro, and Daniel Peters

Main reference J. Fischer, D. Peters, “GLOUDS: Representing tree-like graphs”, Journal of Discrete Algorithms, Vol. 36, pp. 39–49, Elsevier, 2016.

URL <http://dx.doi.org/10.1016/j.jda.2015.10.004>

We plan to combine the recently proposed GLOUDS representation [1] with DSM, a technique used to compress Web and social graphs by exploiting the presence of bicliques and dense subgraphs [2]. Since GLOUDS benefits from a representation with fewer edges per node and DSM reduces the number of edges from $m * n$ to $m + n$ when representing an (m, n) -biclique, we believe the combination can lead to better compression performance than the one obtained with DSM alone, and can offer reasonable edge extraction time.

References

- 1 J. Fischer and D. Peters. GLOUDS: Representing tree-like graphs. J. Discrete Algorithms 36:39–49 (2016).
- 2 C. Hernández, G. Navarro. Compressed representations for web and social graphs. Knowl. Inf. Syst. 40(2):279–313 (2014)

Participants

- Diego Arroyuelo
TU Federico Santa María –
Valparaíso, CL
- Hideo Bannai
Kyushu University –
Fukuoka, JP
- Djamal Belazzougui
CERIST – Algiers, DZ
- Philip Bille
Technical University of Denmark
– Lyngby, DK
- Stefan Böttcher
Universität Paderborn, DE
- Rodrigo Cánovas
University of Montpellier 2, FR
- Patrick Hagge Cording
Technical University of Denmark
– Lyngby, DK
- Héctor Ferrada
University of Helsinki, FI
- Johannes Fischer
TU Dortmund, DE
- Travis Gagie
Universidad Diego Portales, CL
- Adrià Gascón
University of Edinburgh, GB
- Pawel Gawrychowski
University of Wrocław, PL
- Simon Gog
KIT – Karlsruher Institut für
Technologie, DE
- Inge Li Gørtz
Technical University of Denmark
– Lyngby, DK
- Cecilia Hernández Rivas
University of Concepción, CL
- Danny Hucke
Universität Siegen, DE
- Tomohiro I
Kyushu Institute of
Technology, JP
- Shunsuke Inenaga
Kyushu University –
Fukuoka, JP
- Artur Jez
University of Wrocław, PL
- Juha Kärkkäinen
University of Helsinki, FI
- Susana Ladra González
University of A Coruña, ES
- Markus Lohrey
Universität Siegen, DE
- Sebastian Maneth
University of Edinburgh, GB
- Ian Munro
University of Waterloo, CA
- Gonzalo Navarro
University of Chile –
Santiago de Chile, CL
- Yakov Nekrich
University of Waterloo, CA
- Patrick K. Nicholson
Bell Labs – Dublin, IE
- Alberto Ordóñez
University of A Coruña, ES
- Fabian Peternek
University of Edinburgh, GB
- Nicola Prezza
University of Udine, IT
- Rajeev Raman
University of Leicester, GB
- Wojciech Rytter
University of Warsaw, PL
- Hiroshi Sakamoto
Kyushu Institute of Technology –
Fukuoka, JP
- Srinivasa Rao Satti
Seoul National University, KR
- Markus Schmid
Universität Trier, DE
- Manfred Schmidt-Schaus
Goethe-Universität –
Frankfurt a. M., DE
- Rahul Shah
Louisiana State University –
Baton Rouge, US
- Ayumi Shinohara
Tohoku University – Sendai, JP
- Tatiana Starikovskaya
University Paris-Diderot, FR
- Alexander Tiskin
University of Warwick –
Coventry, GB
- Rossano Venturini
University of Pisa, IT



Adaptive Isolation for Predictability and Security

Edited by

Tulika Mitra¹, Jürgen Teich², and Lothar Thiele³

1 National University of Singapore, SG, tulika@comp.nus.edu.sg

2 Friedrich-Alexander-Universität Erlangen-Nürnberg, DE,
teich@informatik.uni-erlangen.de

3 ETH Zürich, CH, thiele@ethz.ch

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 16441 “Adaptive Isolation for Predictability and Security”. Semiconductor technology is at the verge of integrating hundreds of processor cores on a single device. Indeed, affordable multi-processor system-on-a-chip (MPSoC) technology is becoming available. It is already heavily used for acceleration of applications from domains of graphics, gaming (e.g., GPUs) and high performance computing (e.g., Xeon Phi). The potential of MPSoCs is yet to explode for novel application areas of embedded and cyber-physical systems such as the domains of automotive (e.g., driver assistance systems), industrial automation and avionics where non-functional aspects of program execution must be enforceable. Instead of best-effort and average performance, these real-time applications demand timing predictability and/or security levels specifiable on a per-application basis. Therefore the cross-cutting topics of the seminar were methods for temporal and spatial isolation. These methods were discussed for their capabilities to enforce the above non-functional properties without sacrificing any efficiency or resource utilization. To be able to provide isolation instantaneously, e.g., even for just segments of a program under execution, adaptivity is essential at all hardware- and software layers. Support for adaptivity was the second focal aspect of the seminar. Here, virtualization and new adaptive resource reservation protocols were discussed and analyzed for their capabilities to provide application/job-wise predictable program execution qualities on demand at some costs and overheads. If the overhead can be kept low, there is a chance that adaptive isolation, the title of the seminar, may enable the adoption of MPSoC technology for many new application areas of embedded systems.

Seminar October 30–4, 2016 – <http://www.dagstuhl.de/16441>

1998 ACM Subject Classification C.3 Special-Purpose and Application-Based Systems: Real-time and embedded systems

Keywords and phrases Adaptive isolation, Embedded systems, Real-Time systems, Predictability, Security, MPSoC, Parallel computing, Programming models, Timing analysis, Virtualization

Digital Object Identifier 10.4230/DagRep.6.10.120



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

Adaptive Isolation for Predictability and Security, *Dagstuhl Reports*, Vol. 6, Issue 10, pp. 120–153

Editors: Tulika Mitra, Jürgen Teich, and Lothar Thiele



DAGSTUHL
REPORTS

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

1 Executive Summary

Tulika Mitra

Jürgen Teich

Lothar Thiele

License © Creative Commons BY 3.0 Unported license
© Tulika Mitra, Jürgen Teich, and Lothar Thiele

Semiconductor industry has shifted from processor clock speed optimization (having reached its physical limits) to parallel and heterogeneous many-core architectures. Indeed, the continuous technological scaling enables today the integration of hundred and more cores and, thus, enormous parallel processing capabilities. Whereas higher (average) performance has been and still is the major driver for any MPSoC platform design, there is a huge hesitation and fear to install such platforms in embedded systems that require predictable (boundable) guarantees of non-functional properties of execution rather than average properties for a mix of applications. Moreover, it may be observed that in an embedded system, each application running on a platform typically a) requires different qualities to be satisfied. For example, one application might demand for authentication, thus requiring the guarantee of unmodified data and program but have no requirements on speed of execution. Another application might rather require the execution to meet a set of real-time properties such as a deadline or a target data rate. To give an example, consider a driver assistance video processing application in a car that must detect obstacles in front of the car fast enough so to activate the brake system in a timely manner. It must therefore be possible to enforce a set of non-functional qualities of execution on a multi-core platform on a per-application/job basis. b) The above requirements on execution qualities may even change over time or during the program execution of a single application or being dependent on user or environmental settings. For example, one user might not care about sending or distributing personal information over the communication interfaces of a mobile phone whereas another one cares a lot, even in the presence of side channels.

Unfortunately, the way MPSoCs are built and programmed today, the embedded system engineers often experience even worse execution qualities than in the single core case, the reason being the sharing of resources such as cores, buses and/or memory in an unpredictable way. Another obstacle for a successful deployment of multi-core technology in embedded systems is the rather unmanageable complexity. This holds particularly true for the analysis complexity of a system for predictable execution qualities at either compile-time or run-time or using hybrid analysis techniques. The complexity is caused here by an abundant number of resources on the MPSoC and the increasing possibilities of interference created by their concurrent execution and multiple layers of software controlling program executions on a platform. Such layers are often designed for contradictory goals. For example, the power management firmware of an MPSoC may be designed to reduce the energy/power consumption or avoid temperature hot spots. The OS scheduler, on the other hand, may be designed to maximize the average CPU utilization for average performance. Providing tight bounds on execution qualities of individual applications sharing an execution platform is therefore not possible on many MPSoC platforms available today.

One remedy out of this dilemma that has been proposed a long time before the introduction of any MPSoC technology is isolation. With isolation, a set of techniques is subsumed to separate the execution of multiple programs either spatially (by allocating disjoint resources) or temporally (by separating the time intervals shared resources are used). Additionally, in order to provide isolation on demand, there is the need for adaptivity in all hardware as

well as software layers from application program to executing hardware platform. Indeed, adaptivity is considered a key topic in order to reduce or bound execution quality variations actively on a system and in an on-demand manner for the reason to neither overly restrict nor to underutilize available resources.

Adaptive Isolation, the topic of the proposed Dagstuhl seminar, may be seen as a novel and important research topic for providing predictability of not only timing but also security and may be even other properties of execution on a multi-core platform on a per application/job basis while easing and trading off compile-time and run-time complexity.

First, a common understanding of which techniques may be used for isolation including hardware units design, resource reservation protocols, virtualization techniques, and including novel hybrid and dynamic resource assignment techniques were discussed. Second, a very interdisciplinary team of experts including processor designers, OS and compiler specialists, as well as experts for predictability and security analysis were brought together for evaluating these opportunities and presenting novel solutions. The competencies, experiences, and existing solutions of the multiple communities stimulated discussions and co-operations that hopefully will manifest in innovative research directions for enabling predictability on demand on standard embedded MPSoCs.

2 Table of Contents

| | |
|---|-----|
| Executive Summary | |
| <i>Tulika Mitra, Jürgen Teich, and Lothar Thiele</i> | 121 |
| Major Topics Discussed | 125 |
| Adaptive Isolation for Timing Predictability | 125 |
| Isolation and Adaptivity for Security | 125 |
| Cross-Cutting Concerns | 126 |
| Summary of the Presentations | 126 |
| Predictability | 128 |
| Security | 130 |
| Cross-cutting Concerns for Adaptive Isolation | 130 |
| Abstract of Talks | 131 |
| Network-on-Chip-Assisted Adaptive Partitioning and Isolation Technology for “Dynamic” Homogeneous Manycores <i>Davide Bertozzi and Balboni Marco</i> | 131 |
| Use only when you need – Providing adaptive temporal isolation in Cyber-Physical Systems <i>Samarjit Chakraborty</i> | 132 |
| Achieving Timing Predictability by Combining Models <i>Heiko Falk and Arno Luppold</i> | 133 |
| Soteria: Offline Software Protection within Low-cost Embedded Devices <i>Johannes Götzfried</i> | 134 |
| Challenges of Temporal Isolation <i>Gernot Heiser</i> | 134 |
| Predictability in Multicore Systems Using Self-Suspending Tasks <i>Jian-Jia Chen</i> | 135 |
| Software Development for Isolation <i>Tulika Mitra</i> | 135 |
| Time-Based Intrusion Detection in Cyber-Physical Systems <i>Frank Mueller</i> | 136 |
| Adaptive Pipeline for Security in Real Time Systems <i>Sri Parameswaran</i> | 136 |
| Timing Predictability and How to Achieve It <i>Jan Reineke</i> | 137 |
| Connecting the dots – Towards the total automation of embedded systems design (in Java world) <i>Zoran Salcic</i> | 138 |
| Isolation for Security <i>Patrick Schaumont</i> | 138 |

| | |
|--|-----|
| T-CREST: Time-predictable Multi-Core Architecture for Embedded Systems <i>Martin Schoeberl</i> | 139 |
| Adaptive Memory Protection for Many-Core Systems <i>Wolfgang Schröder-Preikschat</i> | 140 |
| Security Issues on the Boundary of Ideal and Real Worlds <i>Takeshi Sugawara</i> | 141 |
| An Introduction to the Seminar <i>Jürgen Teich</i> | 141 |
| Isolation, resource efficiency and covert channels <i>Lothar Thiele</i> | 142 |
| Determinate and Timing-Predictable Concurrency in Reactive Systems – The Synchronous Approach and the SCCharts Language <i>Reinhard von Hanxleden</i> | 142 |
| Hybrid Application Mapping for Dynamic Isolation in Invasive Computing <i>Stefan Wildermann</i> | 143 |
| Timing Verification – Flogging a Dead Horse? <i>Reinhard Wilhelm</i> | 143 |
| Working groups | 144 |
| Runtime Monitoring <i>Felix Freiling</i> | 144 |
| Future of Timing Verification <i>Samarjit Chakraborty</i> | 147 |
| Attack Models <i>Albert Cohen and Karine Heydemann</i> | 148 |
| Synergy between Predictability and Security <i>Frank Mueller</i> | 149 |
| Models of Computation and Programming Languages <i>Reinhard von Hanxleden</i> | 150 |
| Panel Discussion | 152 |
| Acknowledgements | 152 |
| Participants | 153 |

3 Major Topics Discussed

In the following, major topics and questions that were raised and discussed during the seminar, are summarized.

3.1 Adaptive Isolation for Timing Predictability

- New ways to establish isolation by means of hardware and software: Which of the approaches and known concepts for isolation can be used in adaptive scenarios, which rather not?
- Analysis: Statistical vs. hard guarantees? What are limitations of either approach? Can these techniques be reasonably generalized to other architectural elements besides caches?
- Hybrid Analysis and Resource Management: Novel techniques for Mixed Static/Dynamic Resource Assignment and Analysis. Which improvements (e.g., reduced search space vs. less pessimistic bounds) may these techniques deliver and for which set of applications (e.g., periodic streaming, DSP, aperiodic real-time control, mixed critical applications) may these be applied? How may the search space for design decisions regarding resource assignment and scheduling be reduced to a minimum through a characterization of static vs. run-time?
- Online isolation through reconfiguration (e.g., dynamic TDMA adaptation, switching protocols, dynamic schedule adaptation).
- Adaptive hardware architectures (e.g., processor buses with switchable protocols: static priority vs. TDMA depending on workload mix at run-time).
- Utilization and timing analysis for unknown execution time and workload scenarios.
- Cost/Benefit Analysis of adaptive isolation techniques: How much more expensive are adaptive techniques in relation to conventional techniques (Hardware/Software Overheads, adaptation time (e.g., switching times, optimization times, times until stabilization, utilization gains, etc.).
- How can we bound the interference between tasks due to heat transfer?

3.2 Isolation and Adaptivity for Security

- Definition of security in an adaptive MPSoC context. How do security issues change by introducing adaptivity? What is the attackers' model?
- Security bottlenecks of current MPSoC systems with respect to hardware architecture and the possibilities to isolate applications.
- Security requires a root of trust. Security also makes use of isolation. We should reason about secure hand-over in the context of adaptivity. When software modules move from one hardware unit to another one, how are the root of trust and isolation transferred?
- With respect to which properties may security be defined? For example, basic isolation might be defined as a guarantee that no other application may read or write the data of another. For example, a designer or user of an app might require that the data entered or processed to be confidential or request a guarantee that it is unaltered.
- Which techniques must be available at the hardware and software side to enforce certain levels of security on a per-application basis on an MPSoC platform and what is the expected overhead of such techniques?
- May different levels of per-application/job security also be established adaptively?
- Hardware architecture designs for adaptive security.

- Do there exist other levels of security? For example, side channel attacks? Which isolation techniques may be employed on an MPSoC to restrict, prevent, or minimize the chances of attacks, e.g., in terms of resource isolation through resource allocation techniques, encryption on demand on a Network-on-Chip, etc.?
- Is heat transfer a side-channel information leakage source? Can it be a threat to privacy and security? How can we quantify the corresponding effects and what are reasonable countermeasures?

3.3 Cross-Cutting Concerns

From a resource management's point of view, modern embedded system applications come with significant challenges: Highly dynamic usage scenarios as already observable in today's "smart devices" result in a varying number of applications, each with different characteristics, but running concurrently at different points in time on a platform. Obviously, full isolation, avoiding any resource sharing (e.g., by partitioning) is generally too costly or inefficient (utilization). No isolation, on the other hand, will not allow for timing and security properties to hold. From the programmer's point of view, strong isolation and efficient sharing are desired, but they represent two opposing goals.

Traditional techniques to provide a binding or pinning of applications to processors are either applied at design time and result in a static system design. Such a static design may, on the one hand, be too optimistic by assuming that all assigned resources are always available or it may require for over-allocation of cores to compensate for worst-case scenarios.

In this area, cross-cutting techniques such as partitioning, gang scheduling, dynamic resource allocation, virtualization, e.g., real-time scheduling in hypervisors, are opportunities that were discussed for their capability for providing some degree of isolation and capabilities of providing quality on demand per application/job.

Finally, the interaction between security and timing predictability were explored. A malware can compromise a real-time system by making an application miss its deadline and the system should ensure that deadline overruns in the presence of malware be predicted early and remedial actions taken. On the other hand, as the bounds on execution times of an application are known in real-time systems, an execution time outside the bound indicates the possibility of unauthorized code execution and provides an additional mechanism for malware detection. The scheduling and resource allocation should also take into account the trade-off between the timing overheads of security protection mechanism (e.g., encryption cost) leading to increased execution time (and hence difficulty in schedulability) vis-à-vis the need for security isolation.

4 Summary of the Presentations

The presentations in this seminar included state-of-the-art adaptive isolation techniques for both security and predictability. Five breakout sessions covered discussions on the current status, future challenges and research opportunities. A very interdisciplinary team of experts including processor designers, OS and compiler specialists, as well as experts on predictability and security evaluated these opportunities and presented novel solutions. This subsection presents an overview of the topics covered by individual speakers in the seminar. Please refer to the included abstracts to learn more about the individual presentations.

The seminar opened with an introduction by organizer Jürgen Teich (Friedrich-Alexander-Universität Erlangen-Nürnberg). He explained the motivation behind the seminar in the

context of emerging many-core architectures. These architectures can potentially be deployed in embedded systems with strict timing and security guarantee requirements. He presented different definitions of timing predictability and sources of unpredictability such as resource sharing, multi-threading, and power management. He briefly talked about adaptive isolation techniques, such as resource reservation and virtualization, developed in the context of the Invasive Computing (InvasIC) project¹ – a DFG-funded Transregional Collaborative Research Center investigating a novel paradigm for the design and resource-aware programming of future parallel computing systems. He emphasized the similarities between the adaptive isolation techniques for security and timing predictability – a key theme of the seminar.

The introduction was followed by two keynote talks: one on predictability and one on security. The predictability keynote was delivered by Jan Reineke (Universität des Saarlandes). He explained two sources of variation in execution time for software timing analysis: the program input and the micro-architectural state. He raised the key concern that interference due to resource sharing (for L2 cache and memory controller for example) can lead to significant slowdown on multi-core platform compared to single-threaded execution. He stressed the importance of deriving accurate timing models given the lack of information available regarding timing in the underlying architecture. He defined predictability and analyzability as two important but different properties essential towards accurate software timing analysis. Both predictability and analyzability can be enhanced by eliminating stateful micro-architectural components by stateless ones (e.g., replacing caches with scratchpad memory), eliminating interference in shared resources through isolation, and choosing “forgetful” micro-architectural components (e.g., Pseudo LRU replacement policy in place of LRU). In addition, analyzability can be improved if the underlying platform exhibits freedom from timing anomalies (local worst case does not lead to global worst case in systems with timing anomaly) and offers timing compositionality. He presented strictly in-order pipeline processors as an example of such an ideal platform; but the performance impact of such an architecture and its commercial viability remain unknown.

The keynote talk on security was delivered by Patrick Schaumont (Virginia Polytechnic Institute). He motivated the need for secure isolation by introducing a contemporary trusted medical application where privacy/security mechanisms need to be enforced in an end-to-end fashion from tiny micro-controllers (for sensing) to more powerful multi-cores (for gateway device) and finally to servers with complex processors, large memory, and huge storage (for data analytics). He emphasized the key concerns in such platforms, namely, security, safety, and privacy, that demand isolated storage, communication, and execution. The two building blocks of secure computing are the trust boundary and the attacker models that breach the trust boundaries. Isolation is one (but not the only) way to achieve trust by providing confidentiality guarantees in a secure implementation. However, it is important to remember that complete isolation is not a feasible alternative and isolation for security almost always incurs overhead either in terms of area or performance just like predictability. He then presented two examples of isolation for security: SANCUS for lightweight isolation in micro-controllers and SGX for server class isolation. In closing, Patrick mentioned few open challenges such as quantifying security and its resource overhead through well-defined metrics and classifying properties of secure computing in general and secure computer architectures in particular.

¹ <http://www.invasic.de>

4.1 Predictability

The topics covered under adaptive isolation for predictability centered around the future of predictability, design of predictable architectures, providing isolation in general-purpose multi-/many-core architecture, and predictability in reactive systems.

4.1.1 Future of predictability

The talks on timing predictability presented two contrasting views. Reinhard Wilhelm (Universität des Saarlandes) concurred with Jan Reineke’s viewpoint in the keynote that predictability and analyzability are becoming increasingly challenging and even impossible with continuous advances in commercial micro-architectures that harm rather than aid predictability. Architectural complexity leads to analysis complexity. The recipe for success in timing analysis has been abstraction and decomposition. Unfortunately, contemporary processors – even processors supposed to be designed for real-time systems (such as ARM Cortex R5F) – include features (e.g., random replacement caches) that make abstraction and decomposition infeasible. Alternatives to static timing analysis, such as measurement-based methods, do not offer soundness and at the same time suffer from lack of accurate timing models just like static analysis.

In contrast to these views that embedded systems require complete timing predictability, Samarjit Chakraborty (TU München) claimed that in certain applications, such as control systems, it is possible to live with less than total timing predictability. As most controllers exhibit certain degree of “robustness”, the behavior of the controller will not be impacted if some deadlines are missed. Thus, timing analysis, instead of focusing on deadline constraints, should focus on higher-level (control theoretic) goals that better characterize system performance requirements. Achieving this, however, requires quantifying the robustness of the controller to identify the deadlines that are crucial to be satisfied and the ones that can be ignored without any major impact on controller outcome.

4.1.2 Predictable Architecture and Optimizations

Reinhard Wilhelm presented a constructive approach called PROMPT architecture that provides timing isolation for each task when executing multiple tasks on a multi-core architecture. The generic PROMPT architecture is instantiated for each application so as to minimize interferences among the tasks of the application as much as possible. The idea of time predictable architecture was also revisited by Martin Schoeberl (Technical University of Denmark) who presented T-CREST, a time-predictable multi-core architecture for embedded systems. The vision behind T-CREST is to make the worst-case fast and the whole system analyzable rather than make the common case fast as is the conventional wisdom in general-purpose architecture community. The architecture provides constant-time execution of instructions, time-division-multiplexing in the Network-on-Chip (NoC), and software-controlled scratchpad memory for predictability. More importantly, T-CREST provides a complete platform implemented in FPGAs as well as simulator supporting both compiler and analysis tools released under open source BSD license.

Zoran Salcic (University of Auckland) presented an orthogonal solution for timing predictability starting from formal specification of the system in SystemJ, which is based on a formal model of computation. The key feature of SystemJ is Globally Asynchronous Locally Synchronous (GALS) model while incorporating Java for objects and computations allowing SystemJ specification to be executable on any Java processor. He presented an automated design flow that can translate the formal specification to custom NoC-based heterogeneous

multiprocessor platform through design space exploration, optimizations, and scheduling. This is similar in vein to the PROMPT approach, except that software code, schedule and platform instance are all generated automatically in this approach. He concluded his talk by demonstrating an automated bottling machine designed with this model-driven approach.

Heiko Falk (TU Hamburg-Harburg) presents his vision to achieve predictability by combining models during compilation. In current software design practice for real-time systems, the software is designed and optimized for the average-case behavior followed by software timing analysis to ensure that the execution meets deadline constraints. He proposed design of WCET-aware compiler that optimizes software for the worst-case execution time rather than average case. This is achieved by integrating the timing models initially designed for analysis in the compiler itself. Combined with more predictable architectural components, such as scratchpad memory instead of cache, the WCC compiler can provide resource isolation and enables schedulability in some systems that could not meet deadlines under existing software design process.

4.1.3 Isolation for predictability in multi-core

Stefan Wildermann (Universität Erlangen-Nürnberg) followed up from the introduction by Jürgen Teich on achieving adaptive isolation in the context of Invasive Computing. He described a hybrid mapping approach where the solution for each individual task is obtained statically but these individual solutions are put together at runtime through a compositional timing analysis that relies on a composable NoC. The main idea is to carry out performance analysis and design space exploration for individual tasks at design time and identify a set of Pareto-optimal points. At runtime, depending on the scenario, a set of design points (one per task) are identified that satisfy the constraints for all the tasks. The downside of this approach is the huge runtime to choose these design points and may outweigh the benefits of isolation.

Jian-Jia Chen (TU Dortmund) focused on system-level timing analysis in multi-core systems with multiple tasks. Current two-phase analysis approaches find the WCET of each task individually and then compute the worst-case response time (WCRT) of a set of tasks by considering interference from other tasks for shared resources. However, in the presence of shared resources, a task might be suspended from execution when it cannot get immediate access to the resource. This self-suspension of tasks needs to be accounted for in WCRT analysis. But many existing works fail to handle the impact of self-suspension correctly leading to overly optimistic execution time. His talk pointed out the challenges in providing predictability on multi-cores: isolation through time-division-multiplexing introduces unnecessary pessimism and cannot work if the tasks need to share information. On the other hand, with sharing, WCRT analysis and schedulability tests are not well-equipped to handle the interference that need synergy between scheduler design, program analysis, and models of computation.

4.1.4 Reactive Systems

Reinhard von Hanxleden (Universität Kiel) and Albert Cohen (ENS-Paris) discussed predictability in reactive systems. Reinhard von Hanxleden talked about the power of synchronous programming languages such as SCADE and SCCharts. He showed the extensions to these languages that allow deterministic implementation in hardware/software directly from the model. He emphasized the importance of compilation approach on timing predictability, specially in model-to-model mappings. Albert Cohen presented control systems with significant

computations and how to reconcile the computation with the control. A synchronous language like Lustre is extended with isolation control features and ability to safely accommodate some delay in the computation. Similar to Reinhard von Hanxleden's approach, the compiler plays crucial role in mapping the abstract model and real-time scheduling onto multi-core system with simple runtime support for adaptive isolation.

4.2 Security

In his keynote, Patrick Schaumont talked about hardware architectures needed for secure isolation. Johannes Götzfried (Universität Erlangen-Nürnberg) presented Soteria – a lightweight solution for secure isolation in low-cost embedded devices. Soteria can effectively protect the confidentiality and integrity of an application against all kinds of software attacks including attacks from the system level. Soteria achieves this through a simple program-counter based memory access control extension for the TI MSP430 microprocessor with minimal overhead.

Tulika Mitra (National University of Singapore) mentioned the challenges associated with the adoption of secure isolation mechanisms by software developers. She presented an automated approach that, given an Android application, can identify the sensitive code fragments, move them to the secure world (ARM TrustZone), and finally re-factor the original application to establish communication between the normal code fragments and the secure code fragments. This automation takes away the burden of utilizing secure isolation mechanisms by software developers.

Lothar Thiele (ETH Zürich) introduced the possibility of thermal covert channels in multi-core systems. He demonstrated that the on-chip temperature sensors can represent a security breach by allowing otherwise isolated applications running on different cores to communicate and possibility leak sensitive data. A quantification of the covert channel capacity leveraging both theoretical results from information theory and experimental data from modern platforms (such as Android phone) showed sufficient bandwidth for the channel to be useful for information leakage.

Sri Parameswaran (UNSW Sydney) presented an online monitoring technique to detect and recover from hardware Trojans in pipelined multiprocessor system-on-chip devices. The system adapts and randomizes to provide security. Takeshi Sugawara (Mitsubishi, Kanagawa) stressed the importance of assumptions (model abstractions) in security. In the context of side-channel attacks and LSI reverse engineering, he showed how the countermeasures are constructed and how their assumptions are falsified. He also presented static isolation based on domain-specific coprocessors.

4.3 Cross-cutting Concerns for Adaptive Isolation

Adaptive isolation mechanisms that can be employed for both security and predictability, as well as the synergy and conflict between security and predictability featured prominently and repeatedly in the seminar.

Gernot Heiser (UNSW Sydney) pointed out the challenges towards isolation from both security and predictability perspective. He opined that spatial isolation is relatively easy to achieve, both in single-core and multi-core settings, given the support from both hardware and software. He cited seL4 micro-kernel as an example to illustrate his point. He, however, re-iterated (just like Reinhard Wilhelm and Jan Reineke) that temporal isolation is much harder especially in the presence of complex, unpredictable hardware. The instruction-set

architecture (ISA) no longer provides a guaranteed contract between hardware and software for either timeliness or security (for example, hidden states and timing channels). He called on the architects to extend the ISA so that timing effects become visible and hardware provides mechanisms to partition or flush shared states with bounded latency so as to provide isolation.

Davide Bertozzi (Università di Ferrara) focused on NoC to provide adaptive isolation in many-core architectures. He had a different opinion from Gernot Heiser regarding spatial isolation and showed that current many-core accelerator architectures are at odds with spatial-division multiplexing. For example, the traffic generated by different applications may collide in the NoC when NoC paths are shared between nodes assigned to different applications even if each core is allocated to only a single application exclusively. He presented routing restrictions as an approach towards partitioning the resources among different applications; but this leads to additional challenges in mapping as well as reconfigurability and adaptivity of the partitions.

Wolfgang Schröder-Preikschat (Universität Erlangen-Nürnberg) presented isolation in memory to protect against unintentional programming errors as well as attacks from malicious programs/processes. While existing memory-management units provide protection, they harm time predictability. There are scenarios where ubiquitous memory protection is unnecessary and increases uncertainty for some time, but is required at other points during the run time of a system. He proposed adaptive memory protection as a solution, where the protection state of applications can change over time. It allows the combination of benefits of both worlds: security when memory protection is needed and increased performance and predictability once security is superfluous.

Sibin Mohan (University of Illinois at Urbana-Champaign) expounded on the interaction between predictability and security. He alerted the audience to the challenges of real-time systems running in insecure world. Real-time systems demand predictability; but predictability actually enables the attackers to precisely reconstruct the execution behavior of the system. In particular, he showed how information about the behavior of real-time systems (e.g., schedule) can be leaked by adversaries and presented techniques to deter such attacks. On the other hand, sometimes it is possible to use the predictable behavioral properties of real-time systems to actually detect intrusion almost as soon as they occur. Frank Mueller (North Carolina State University) also exploited the synergy between predictability and security. His approach utilizes the timing bounds obtained for different code fragments of a program during static timing analysis. At runtime, if the execution time of a code fragment falls outside its pre-determined bounds, the system flags an intrusion

5 Abstract of Talks

5.1 Network-on-Chip-Assisted Adaptive Partitioning and Isolation Technology for “Dynamic” Homogeneous Manycores

Davide Bertozzi (Università di Ferrara, IT) and Balboni Marco

License © Creative Commons BY 3.0 Unported license
© Davide Bertozzi and Balboni Marco

Joint work of Davide Bertozzi, Marco Balboni, Giorgos Dimitrakopoulos, José Flich

The software parallelism is not keeping up with hardware parallelism, therefore the problem of efficiently exploiting large array fabrics of homogeneous processing cores will soon come

to the forefront. Multi-programmed mixed-criticality workloads are the straightforward solution to this problem, although they raise a number of practical issues ranging from system composability techniques for easier verification, to performance predictability and/or security. This talk presents a systematic approach to these issues through an adaptive partitioning and isolation technology for manycore computing fabrics having its key enabler in the reconfigurable features of the on-chip interconnection network. The technology relies on two main pillars. First, a hierarchy of partition types enables to properly sandbox applications with controlled degrees of interactions and/or dependencies (if at all allowed). Second, fine-grained adaptivity of the system configuration to the workload is implemented with NoC assistance for the sake of power-efficient resource management at any given point in time. It follows from this a “design-for-partitioning” philosophy that is at the core of future dynamic hardware platforms, and that will shape their architectures from the ground up.

5.2 Use only when you need – Providing adaptive temporal isolation in Cyber-Physical Systems

Samarjit Chakraborty (TU München, DE)

License © Creative Commons BY 3.0 Unported license
© Samarjit Chakraborty

Joint work of Samarjit Chakraborty, Alejandro Masrur, Ansuman Banerjee, Anuradha M. Annaswamy, Jian-Jia Chen, Dip Goswami, Harald Voit, Reinhard Schneider

Main reference A. Masrur, D. Goswami, S. Chakraborty, J.-J. Chen, A. Annaswamy, A. Banerjee, “Timing analysis of cyber-physical applications for hybrid communication protocols”, in Proc. of the Conf. on Design, Automation and Test in Europe (DATE 2012), pp. 1233–1238, IEEE, 2012.

URL <http://dx.doi.org/10.1109/DATE.2012.6176681>

Many embedded control systems have distributed implementations, in which sensor values and control signals have to be communicated over shared communication buses. The participants sharing the bus along with the bus protocol being used determine the delay suffered by the control signals, which in turn affect stability and control performance. Two broad classes of communication protocols exist, which are based on either the time-triggered or the event-triggered paradigms. The former ensures strict temporal isolation between messages and results in more deterministic communication. Hence, it is easier to use when guarantees on stability and control performance are required. The latter does not provide temporal isolation between messages, but has several advantages like better bus utilization and easier extensibility. This has also resulted in hybrid protocols that combine the event- and time-triggered paradigms. However, there has been little work on how to exploit such hybrid protocols when designing control algorithms, in order to utilize the benefits of both the communication paradigms. In this talk we will discuss this problem and propose some promising research directions that involve adaptively providing temporal isolation on a when-needed basis. This brings up a number of challenges both in the areas of control theory, and also in timing analysis.

References

- 1 Harald Voit, Anuradha M. Annaswamy, Reinhard Schneider, Dip Goswami, Samarjit Chakraborty. *Adaptive switching controllers for systems with hybrid communication protocols*. American Control Conference (ACC) 2012
- 2 Harald Voit, Anuradha Annaswamy, Reinhard Schneider, Dip Goswami, Samarjit Chakraborty. *Adaptive switching controllers for tracking with hybrid communication protocols*. 51th IEEE Conference on Decision and Control (CDC) 2012

- 3 Alejandro Masrur, Dip Goswami, Samarjit Chakraborty, Jian-Jia Chen, Anuradha Anaswamy, Ansuman Banerjee. *Timing analysis of cyber-physical applications for hybrid communication protocols*. Design, Automation & Test in Europe Conference (DATE) 2012
- 4 Dip Goswami, Reinhard Schneider, Samarjit Chakraborty. *Re-engineering cyber-physical control applications for hybrid communication protocols*. Design, Automation & Test in Europe Conference (DATE) 2011

5.3 Achieving Timing Predictability by Combining Models

Heiko Falk (TU Hamburg-Harburg, DE) and Arno Luppold

License © Creative Commons BY 3.0 Unported license
© Heiko Falk and Arno Luppold

Main reference A. Luppold, H. Falk, “Code Optimization of Periodic Preemptive Hard Real-Time Multitasking Systems”, in Proc. of the 18th Int’l Symposium on Real-Time Distributed Computing (ISORC 2015), pp. 35–42, IEEE, 2015.

URL <http://dx.doi.org/10.1109/ISORC.2015.8>

During the design of embedded software, compilers play an important role, since the machine code generated by them directly influences criteria like, e.g., execution times, timing predictability or energy. Particularly, compiler optimizations could be beneficial to improve such criteria systematically.

The discussions during this seminar revealed that both the predictability and the security community lack suitable models and that, if models are available, they are often used in the form of black boxes. This presentation intends to show what can be done within a compiler when combining models that are usually used by different communities.

By coupling a compiler with a static timing analyzer, a formal WCET timing model based on micro-architectural features was integrated into the compilation flow. Next, this low-level hardware model is combined with a code-level control flow model that allows for the systematic optimization of WCETs by the compiler. Finally, task set-level models from the scheduling theory community are integrated into the optimization flow.

By means of a Scratchpad Memory (SPM) allocation, this presentation aims to show how complete multi-task sets can finally be optimized for timing predictability. Due to their timing predictability, SPMs are useful to achieve isolation between concurrent software tasks. By combining all these various models into the compiler’s optimization process, we are able to achieve predictability and inter-task isolation by controlling resource use statically at compile time for entire multi-task systems.

In the future, it would be worthwhile to investigate in how far the memory-related isolation achieved by our existing WCET-oriented optimizations are useful for security. Furthermore, a tight(er) connection between compilers and operating systems might be useful for more efficient and effective resource allocation decisions at runtime in order to finally achieve adaptive isolation.

References

- 1 Arno Luppold, Heiko Falk. *Code Optimization of Periodic Preemptive Hard Real-Time Multitasking Systems*. In Proceedings of the 18th International Symposium on Real-Time Distributed Computing (ISORC), Auckland / New Zealand, April 2015

5.4 Soteria: Offline Software Protection within Low-cost Embedded Devices

Johannes Götzfried (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Johannes Götzfried

Joint work of Johannes Götzfried, Tilo Müller, Ruan de Clercq, Pieter Maene, Felix C. Freiling, Ingrid Verbauwhede

Main reference J. Götzfried, T. Müller, R. de Clercq, P. Maene, F. C. Freiling, I. Verbauwhede, “Soteria: Offline Software Protection within Low-cost Embedded Devices”, in Proc. of the 31st Annual Computer Security Applications Conf. (ACSAC 2015), pp. 241–250, ACM, 2015.

URL <http://dx.doi.org/10.1145/2818000.2856129>

Protecting the intellectual property of software that is distributed to third-party devices which are not under full control of the software author is difficult to achieve on commodity hardware today. Modern techniques of reverse engineering such as static and dynamic program analysis with system privileges are increasingly powerful, and despite possibilities of encryption, software eventually needs to be processed in clear by the CPU. To anyhow be able to protect software on these devices, a small part of the hardware must be considered trusted. In the past, general purpose trusted computing bases added to desktop computers resulted in costly and rather heavyweight solutions. In contrast, we present Soteria, a lightweight solution for low-cost embedded systems. At its heart, Soteria is a program-counter based memory access control extension for the TI MSP430 microprocessor. Based on our open implementation of Soteria as an openMSP430 extension, and our FPGA-based evaluation, we show that the proposed solution has a minimal performance, size and cost overhead while effectively protecting the confidentiality and integrity of an application’s code against all kinds of software attacks including attacks from the system level.

5.5 Challenges of Temporal Isolation

Gernot Heiser (UNSW – Sydney, AU)

License © Creative Commons BY 3.0 Unported license
© Gernot Heiser

Joint work of Gernot Heiser, Anna Lyons, Thomas Sewell, Felix Kam, Qian Ge, Yuval Yarom

Main reference Q. Ge, Y. Yarom, G. Heiser, “Do Hardware Cache Flushing Operations Actually Meet Our Expectations?”, arXiv:1612.04474v3 [cs.CR], 2016.

URL <https://arxiv.org/abs/1612.04474v3>

Spatial isolation is well-supported by present hardware and software, e.g. the seL4 microkernel has been proved to support spatial isolation, including the absence of covert storage channels. While the formal arguments about seL4 presently only apply to a single-core version, the extension its functional verification to multicore hardware is in progress, and unlikely to produce issues in terms of spatial isolation.

In contrast, temporal isolation is not only harder to verify, hardware is becoming less predictable, thanks to an increasing number of performance-enhancement tricks, generally based on some form of caching and dynamic scheduling of resources. This makes it increasingly difficult, and in cases impossible, to bound and control non-determinism.

I argue that computer architects have essentially abandoned the instruction-set architecture (ISA) as the contract between hardware and software: by just referring to the ISA, it is impossible to guarantee safety (timeliness) and security (absence of timing channels).

I argue further that it is hopeless to address this problem unless architects agree to a usable contract, i.e. extend the ISA so that timing effects become visible (and thus analysable) or controllable.

In particular, there must be time bounds on all operations. In practice, bounding each individual operation (instruction) may not be enough, as this will lead to massively pessimistic bounds. As future hardware will never be fully utilisable (eg one cannot run all cores because they will overheat), this pessimism may be tolerable in many cases. In others, enough information must be available so that it is at least possible to obtain realistic bounds on the execution time of groups of operations, giving software the opportunity to re-introduce determinism at a higher level.

Examples of this are variations produced by shared state such as various forms of caches and interconnects, which produce variations in execution time that break isolation. Establishing safety requires the ability to bound variations. Establishing security is harder, as it requires establishing determinism, at least at some course granularity. This is possible as long as the hardware provides mechanisms to either partition or flush (with bounded latency) any such shared state.

5.6 Predictability in Multicore Systems Using Self-Suspending Tasks

Jian-Jia Chen (TU Dortmund, DE)

License © Creative Commons BY 3.0 Unported license
© Jian-Jia Chen

In general computing systems, a job (process/task) may suspend itself whilst it is waiting for some activity to complete. With the presence of self-suspension, the typical critical instant theorem cannot be directly applied. However, such suspending behavior is in general unavoidable unless the executions are isolated. In this talk, I present a short overview of typical schedulability tests, explain our observations why suspension is important to account for the impact of shared resources, and provide a brief overview of recent developments with regard to the schedulability tests.

5.7 Software Development for Isolation

Tulika Mitra (National University of Singapore, SG)

License © Creative Commons BY 3.0 Unported license
© Tulika Mitra

Joint work of Tulika Mitra, Konstantin Rubinov, Lucia Rosculete, Abhik Roychoudhury
Main reference K. Rubinov, L. Rosculete, T. Mitra, A. Roychoudhury, “Automated Partitioning of Android Applications for Trusted Execution Environments”, in Proc. of the 38th Int’l Conf. on Software Engineering (ICSE’16), pp. 923–934, ACM, 2016.
URL <http://dx.doi.org/10.1145/2884781.2884817>

The co-existence of critical and non-critical applications on computing devices is becoming commonplace. The sensitive segments of a critical application should be executed in isolation on Trusted Execution Environments (TEE) so that the associated code, data, and their execution can be protected from malicious applications both for security and timing predictability. TEE is supported by different technologies and platforms, such as ARM Trustzone, that allow logical separation of secure and normal worlds. However, software development on such platforms to take advantage of the hardware support for isolation remain challenging resulting in slow adoption of isolation techniques at application level. We develop an automated approach to help application developers adopt hardware-enforced secure technology for isolation by retrofitting original applications to protect sensitive data. Our approach automatically partitions critical Android applications into client code to be

run in the normal world and TEE code encapsulating the handling of confidential data to be run in the secure world. We further reduce the overhead due to transitions between the two worlds. The advantage of our proposed solution is evidenced by efficient automated partitioning of real-world Android applications to protect sensitive code/data.

5.8 Time-Based Intrusion Detection in Cyber-Physical Systems

Frank Mueller (North Carolina State University – Raleigh, US)

License © Creative Commons BY 3.0 Unported license
© Frank Mueller

Joint work of Christopher Zimmer, Balasubramanya Bhat, Frank Mueller, Sabin Mohan

Main reference C. Zimmer, B. Bhat, F. Mueller, S. Mohan, “Time-based intrusion detection in cyber-physical systems”, in Proceedings of the 1st ACM/IEEE Int’l Conf. on Cyber-Physical Systems (ICCPS’10), pp. 109–118, ACM, 2010.

URL <http://dx.doi.org/10.1145/1795194.1795210>

Security in real-time cyber-physical systems (CPS) has been an afterthought, even though such systems are networked. We present three mechanisms for time-based intrusion detection exploiting information obtained by static timing analysis. For real-time CPS systems, timing bounds on code sections are readily available as they are calculated during schedulability analysis. We demonstrate how checks of micro-timings at multiple granularities of code uncover intrusions (1) in a self-checking manner by the application and (2) through the operating system scheduler, which has never been done before.

5.9 Adaptive Pipeline for Security in Real Time Systems

Sri Parameswaran (UNSW – Sydney, AU)

License © Creative Commons BY 3.0 Unported license
© Sri Parameswaran

Joint work of Amin Malekpour, Sri Parameswaran, Roshan Ragel

Hardware Trojans are employed by adversaries to either leak information or to prevent computation deliberately by inserting alterations at design time. Hardware Trojans compromise the operation of systems, reducing the trust placed in any manufactured hardware, as well as any software executing upon that hardware. A Trojan can be always ON or be triggered by a certain condition either external or internal. Even before the manufacturing process, intellectual property (3PIPs) cores supplied by third-party vendors as well as electronic design automation (EDA) tools (developed by various companies) could well make the in-house design process of ICs vulnerable. During the typical development cycle of an IC, each party associated with design, manufacturing and distribution of an IC can be a potential adversary, who could well insert undesired malicious modifications into the IC. Therefore, either ensuring that the ICs are free of hardware Trojans or mitigating their harmful impact is important. Most existing countermeasures focus on the difficult task of detecting and preventing hardware Trojans. Although Trojan identification before ICs are deployed in the system can be beneficial, the proposed techniques for detection cannot guarantee detection of all types and sizes of Trojans. We aim to apply online monitoring notion to a Pipelined Multiprocessor System-on-Chip (PMPSoC), which enables the system to work safely in the presence of Trojans while utilizing shelf commercial processing elements (3PIPs). The system adapts and randomizes to provide security. Our proposed online monitoring would facilitate the detection/recovery of/from hardware Trojan attacks, albeit with some overheads. Our system is implemented as PMPSoC architecture and uses a diverse set of 3PIPs.

5.10 Timing Predictability and How to Achieve It

Jan Reineke (*Universität des Saarlandes, DE*)

License © Creative Commons BY 3.0 Unported license
© Jan Reineke

Main reference S. Hahn, J. Reineke, R. Wilhelm, “Toward Compact Abstractions for Processor Pipelines”, in Proc. of the Correct System Design Symposium in Honor of Ernst-Rüdiger Olderog on the Occasion of His 60th Birthday, LNCS, Vol. 9360, pp. 205–220, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-23506-6_14

For hard real-time systems, timeliness of operations has to be guaranteed. Static timing analysis is therefore employed to compute upper bounds on the execution times of a program. Analysis results at high precision are required to avoid over-provisioning of resources.

In the first part of the talk, I stress the need for faithful microarchitectural models. Without such models no reliable predictions about a program’s future execution times can be made. Unfortunately, models at the level of detailed required for timing analysis are rarely available.

In the second part of the talk, I discuss the notions of timing predictability and analyzability. Timing predictability is related to the range of possible execution times of a program under different conditions, such as different initial hardware states or different amounts of interference generated by co-running tasks on other processor cores. Modern microarchitectural features such as deep pipelines, complex memory hierarchies, and shared resources in multi or many cores generally decrease predictability.

I discuss three approaches to increase predictability:

1. Eliminating stateful components: e.g. by replacing caches by scratchpad memories, or an out-of-order architecture by a VLIW architecture. The challenge then is the efficient static allocation of resources.
2. Eliminating interference: this is achieved by partitioning shared resources in time and/or space. The challenge is the efficient partitioning of the resources.
3. Choosing “forgetful” components, i.e., components whose behavior is relatively insensitive to its initial state. For caches we know that LRU replacement in this regard. For other microarchitectural components, our understanding is less developed.

Timing analyzability is concerned with analysis efficiency. Analyzability can be improved by the same three approaches that increase predictability:

1. Eliminating stateful resources results in fewer hardware states that timing analysis needs to account for.
2. Eliminating interference allows timing analysis to safely ignore the behavior of co-running tasks.
3. Different initial states will quickly converge during analysis for forgetful components.

While the three approaches discussed above are beneficial for both predictability and analyzability, two properties of timing models are related primarily to analyzability:

1. Freedom from timing anomalies, and
2. Timing compositionality

Both properties enable the implicit and thus efficient treatment of large sets of hardware states during timing analysis. I show that even models of simple in-order processors are neither free from timing anomalies nor timing compositional. Finally, I sketch “strictly in-order pipelines”, which are provably free from timing anomalies and timing compositional.

5.11 Connecting the dots – Towards the total automation of embedded systems design (in Java world)

Zoran Salcic (University of Auckland, NZ)

License  Creative Commons BY 3.0 Unported license
© Zoran Salcic

Java, although used in large number of embedded systems, still has not found its proper place in research community. By extending Java with GALS abstractions and formal model of concurrency and reactivity, we give it a new life. SystemJ language, initially aimed at general concurrent and distributed systems has found its way to embedded real-time world to become a language with which a design begins and goes through various transformations until it finds a suitable/customised multicore platform for its execution. We will talk about how the dots are connected, the central role of a formal representation of SystemJ program in all phases and variations of design process. Multi-dimensional research has been developed related to generation of efficient code that runs on a time-predictable multi-core platform, satisfies timing constraints proven via static analysis and allows generation of the execution platform suitable for further requirements such as fault-tolerance, isolation of software behaviours using spatial and temporal methods within the platform's NoC, resource-aware program execution etc.

5.12 Isolation for Security

Patrick Schaumont (Virginia Polytechnic Institute – Blacksburg, US)

License  Creative Commons BY 3.0 Unported license
© Patrick Schaumont

Modern information infrastructure is very heterogeneous, with the Internet of Things on the one end, and the Cloud on the other. Computing capabilities, storage, and computing performance vary by orders of magnitude as one moves from the IoT to the cloud. I used the example of implantable/wearable medical devices to illustrate this point, and to address the security concerns that occur within such information infrastructure [1, 2]. In particular, there are requirements for information security, safety, and privacy. These requirements affect the entire information chain from implanted device up to the cloud server.

Hence, when considering 'isolation for security', it is vital to do this within the proper architectural context. Second, the architecture has significant impact on the implementation of security. Constrained implementations, found near the outer periphery in the internet of things, use simple micro-controllers, simple cryptography, and static secrets. High-end implementations, found in the cloud, use advanced processors, complex public-key crypto, and ephemeral secret. Moreover, many of the high-end architectures have isolation naturally build-in.

A central concept in the design of secure architectures is that of trust. A computer system is trusted when it behaves as expected. A computer system that is not trusted has unknown behavior – that is, it may work as we expect it should, or it may not. We just don't know. The boundary between the trusted part and the non-trusted part of a computer system is the trust boundary.

A closely related concept is that of the attacker model [3]. An attacker model enumerates the mechanisms through which the adversary will try to breach the trust boundary. In computer systems, this can be done in various ways. The I/O attacker model assumes an

attacker who can manipulate the input of a system in order to cause an internal exception and take control. The Machine code attacker model assumes an attacker who coexists in the same memory space as a secure task, thereby introducing the requirement to enforce strict isolation between the secure task and the rest of the system. The Hardware attacker model represents the strongest attack and assumes an attacker who has (to some extent) physical control over the computer architecture.

During the talk, I discussed two examples of computer systems that are able to handle the Machine code attacker model, including SGX from Intel [5, 6] and SANCUS, developed at KU Leuven [4].

Some final open issues are (a) how security can be quantified; (b) what metrics would be suitable to describe secure computing and (c) what are the orthogonal properties of secure computing (next to isolation).

References

- 1 Michael Rushanan, Aviel D. Rubin, Denis Foo Kune, Colleen M. Swanson: *SoK: Security and Privacy in Implantable Medical Devices and Body Area Networks*. IEEE Symposium on Security and Privacy 2014: 524–539.
- 2 Wayne Bursleson, Shane S. Clark, Benjamin Ransford, Kevin Fu: *Design challenges for secure implantable medical devices*. DAC 2012: 12–17.
- 3 Frank Piessens, Ingrid Verbauwhede: *Software security: Vulnerabilities and countermeasures for two attacker models*. DATE 2016: 990–999.
- 4 Job Noorman, Pieter Agten, Wilfried Daniels, Raoul Strackx, Anthony Van Herrewege, Christophe Huygens, Bart Preneel, Ingrid Verbauwhede, Frank Piessens: *Sancus: Low-cost Trustworthy Extensible Networked Devices with a Zero-software Trusted Computing Base*. USENIX Security Symposium 2013: 479–494.
- 5 Victor Costan, Srinivas Devadas: *Intel SGX Explained*. IACR Cryptology ePrint Archive 2016: 86 (2016).
- 6 Ittai Anati, Shay Gueron, Simon Johnson, Vincent Scarlata: *Innovative Technology for CPU Based Attestation and Sealing*. Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy, HASP 2013.

5.13 T-CREST: Time-predictable Multi-Core Architecture for Embedded Systems

Martin Schoeberl (Technical University of Denmark – Lyngby, DK)

License © Creative Commons BY 3.0 Unported license
© Martin Schoeberl

Main reference M. Schoeberl, S. Abbaspour, B. Akesson, N. Audsley, R. Capasso, J. Garside, K. Goossens, S. Goossens, S. Hansen, R. Heckmann, S. Hepp, B. Huber, A. Jordan, E. Kasapaki, J. Knoop, Y. Li, D. Prokesch, W. Puffitsch, P. Puschner, A. Rocha, C. Silva, J. Sparsø, A. Tocchi, “T-CREST: Time-predictable multi-core architecture for embedded systems”, *Journal of Systems Architecture*, Vol. 61(9), pp. 449–471, Springer, 2015.

URL <http://dx.doi.org/10.1016/j.sysarc.2015.04.002>

Real-time systems need time-predictable platforms to allow static analysis of the worst-case execution time (WCET). Standard multi-core processors are optimized for the average case and are hardly analyzable. Within the T-CREST project we propose novel solutions for time-predictable multi-core architectures that are optimized for the WCET instead of the average-case execution time. The resulting time-predictable resources (processors, interconnect, memory arbiter, and memory controller) and tools (compiler, WCET analysis)

are designed to ease WCET analysis and to optimize WCET performance. Compared to other processors the WCET performance is outstanding.

The T-CREST project is the result of a collaborative research and development project executed by eight partners from academia and industry. The European Commission funded T-CREST.

5.14 Adaptive Memory Protection for Many-Core Systems

Wolfgang Schröder-Preikschat (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Wolfgang Schröder-Preikschat

Joint work of Gabor Drescher, Wolfgang Schröder-Preikschat

Memory protection based on MMUs or MPUs is widely applied in all areas of computing from mobile devices to HPC. It is a basic building block to provide security between the OS kernel and applications but also among different applications. However, unprotected execution is also generally possible and typically exerted in resource-restricted environments, for instance embedded systems. Both variants have their advantages and disadvantages. While memory protection ensures safety and security in the face of arbitrary programs, it is also costly to manage. Updates of multi-level page-table data structures, TLB invalidations via inter processor interrupts and page faults on memory accesses are significant sources of unpredictability.

State of the art operating systems statically determine which applications or application's modules run under memory protection and which do not. This assignment of protection does not change at run time. This means, software is either restricted to solely access its own memory regions and this is enforced by hardware mechanisms, or it may access all memory regions freely.

There are scenarios where ubiquitous memory protection is unnecessary and increases uncertainty for some time, but is required at other points during the run time of a system. The simplest example is the execution of a single application, kernel regions need to be protected but otherwise the application may freely access all available memory without error. Once another application is started, both may need confinement. In the case of applications written in a type-safe language, memory-protection overheads are also wasteful, as the application cannot perform arbitrary pointer arithmetic. Another scenario may be real-time applications where time predictability may be of higher interest than security. Finally, applications of the same vendor may trust each other but mistrust software of different origin. Memory protection may be superfluous in this scenario until foreign software is started on the system.

This talk presents an adaptive memory-protection system that is capable of dynamically changing the protection state of applications from protected to unprotected and back again. This adaptability applies at run time to parallel applications utilizing dynamic memory allocation. It allows the combination of the benefits of both worlds: security when memory protection is needed and increased performance and predictability once security is superfluous. Evaluation results for up to 64 cores on a contemporary x86 64 bit server show reduced time complexity of relevant system services from linear to constant time in the unprotected case. Unprotected applications benefit from faster system calls, starting at a 3.5 times speedup. Furthermore, it can be shown that unpredictable running times and system call latencies can be reduced. Nevertheless, the OS maintains the ability to confine applications at any time.

5.15 Security Issues on the Boundary of Ideal and Real Worlds

Takeshi Sugawara (Mitsubishi – Kanagawa, JP)

License © Creative Commons BY 3.0 Unported license
© Takeshi Sugawara

Joint work of Takeshi Sugawara, Daisuke Suzuki, Minoru Saeki, Ryoichi Fujii, Shigeaki Tawa, Ryohei Hori, Mitsuru Shiozaki, Takeshi Fujino

Main reference T. Sugawara, D. Suzuki, R. Fujii, S. Tawa, R. Hori, M. Shiozaki, T. Fujino, “Reversing Stealthy Dopant-Level Circuits”, in Proc. of the Workshop on Cryptographic Hardware and Embedded Systems (CHES 2014), Journal of Cryptographic Engineering, Vol. 5(2), pp. 85–94, Springer, 2015.

URL <http://dx.doi.org/10.1007/s13389-015-0102-5>

Main reference T. Sugawara, D. Suzuki, M. Saeki, M. Shiozaki, T. Fujino, “On Measurable Side-Channel Leaks Inside ASIC Design Primitives”, in Proc. of the Workshop on Cryptographic Hardware and Embedded Systems (CHES 2013), Journal of Cryptographic Engineering, Vol. 4(1), pp. 59–73, Springer, 2014.

URL <http://dx.doi.org/10.1007/s13389-014-0078-6>

Assumption (i.e., model, abstraction) is essential in security because it is the interface between theorists and experimentalists. Good assumption can be easily verified by experimentalists. In the talk, my recent results on the research of side-channel attack and LSI reverse engineering are briefly introduced in order to show examples how the countermeasures are constructed and how their assumptions are falsified. Finally, static isolation based on domain-specific co-processors, that is being common in industry is explained.

5.16 An Introduction to the Seminar

Jürgen Teich (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Jürgen Teich

Presented is an introduction and motivation of the topic of this Dagstuhl Seminar: Adaptive isolation of applications on Multi-Core Systems in order to provide, improve or enforce timeliness of computations as well as security on demand.

First, different definitions on timing predictability are revisited and restriction of input spaces as well as isolation of resources discussed for in improving timing predictability or allowing analyzability at all. Subsequently, major sources of unpredictability are summarized, including sharing of resources, multi-threading, and power management techniques as used today.

For isolation, resource reservation protocols, virtualization techniques and invasive computing, a new paradigm for parallel multi-core computing in which applications “invade” resources on demand in order to restrict the interference with other applications is introduced.

It is shown that for many applications such as stream processing, a formal timing analysis is possible. Also, the variation of execution time may be greatly reduced through the isolation created by invading isolated rather than sharing. Finally, invasive computing may also be used to virtualize a multi-core platform and provide secure islands on demand.

Conclusions are given to point out similarities and differences between isolation techniques for time predictability and properties important in the domain of security.

5.17 Isolation, resource efficiency and covert channels

Lothar Thiele (ETH Zürich, CH)

License © Creative Commons BY 3.0 Unported license
© Lothar Thiele

Joint work of Lothar Thiele, Miedl Philipp

Modern multicore processors feature easily accessible temperature sensors that provide useful information for dynamic thermal management. These sensors were recently shown to be a potential security threat, since otherwise isolated applications can exploit them to establish a thermal covert channel and leak restricted information. Previous research showed experiments that document the feasibility of (lowrate) communication over this channel, but did not further analyze its fundamental characteristics. For this reason, the important questions of quantifying the channel capacity and achievable rates remain unanswered.

To address these questions, we devise and exploit a new methodology that leverages both theoretical results from information theory and experimental data to study these thermal covert channels on modern multicores. We use spectral techniques to analyze data from two representative platforms and estimate the capacity of the channels from a source application to temperature sensors on the same or different cores. We estimate the capacity to be in the order of 300 bits per second (bps) for the same-core channel, i.e., when reading the temperature on the same core where the source application runs, and in the order of 50 bps for the 1-hop channel, i.e., when reading the temperature of the core physically next to the one where the source application runs. Moreover, we show a communication scheme that achieves rates of more than 45 bps on the same-core channel and more than 5 bps on the 1-hop channel, with less than 1% error probability. The highest rate shown in previous work was 1.33 bps on the 1-hop channel with 11% error probability.

5.18 Determinate and Timing-Predictable Concurrency in Reactive Systems – The Synchronous Approach and the SCCharts Language

Reinhard von Hanxleden (Universität Kiel, DE)

License © Creative Commons BY 3.0 Unported license
© Reinhard von Hanxleden

Joint work of Joaquin Aguado, David Broman, Björn Duderstadt, Insa Fuhrmann, Reinhard von Hanxleden, Michael Mendler, Steven Loftus-Mercer, Christian Motika, Owen O’Brien, Partha Roop, Steven Smyth, Alexander Schulz-Rosengarten, KIELER and ELK teams

Main reference R. von Hanxleden, B. Duderstadt, C. Motika, S. Smyth, M. Mendler, J. Aguado, S. Mercer, O. O’Brien, “SCCharts: sequentially constructive statecharts for safety-critical applications: HW/SW-synthesis for a conservative extension of synchronous statecharts”, in Proc. of the 35th ACM SIGPLAN Conf. on Prog. Lang. Design and Implementation (PLDI’14), pp. 372–383, ACM, 2014.

URL <http://dx.doi.org/10.1145/2594291.2594310>

Synchronous programming languages are well established for programming safety-critical reactive systems that demand determinate behavior and predictable timing. One commercially successful example is the graphical modeling language provided by the Safety Critical Application Development Environment (SCADE), which is used for e.g. flight control design. Another, more recently developed synchronous language are SCCharts, a statechart variant that extends classical synchronous programming with a more liberal, but yet determinate scheduling regime for shared variables. SCCharts can be compiled into both software and hardware, with a sequence of structural model-to-model transformations that allow to map the temporal behavior back to the model. The presentation emphasizes that the compilation approach has a high influence on timing predictability.

5.19 Hybrid Application Mapping for Dynamic Isolation in Invasive Computing

Stefan Wildermann (Universität Erlangen-Nürnberg, DE)

License © Creative Commons BY 3.0 Unported license
© Stefan Wildermann

Joint work of Andreas Weichslgartner, Deepak Gangadharan, Stefan Wildermann, Michael Glaß, Jürgen Teich
Main reference A. Weichslgartner, D. Gangadharan, S. Wildermann, M. Glaß, J. Teich, “DAARM: Design-time application analysis and run-time mapping for predictable execution in many-core systems”, in Proc. of the Int’l Conf. on Hardware/Software Codesign and System Synthesis (CODES’14), pp. 34:1–34:10, ACM, 2014.

URL <http://dx.doi.org/10.1145/2656075.2656083>

Multi-Processor Systems-on-a-Chip (MPSoCs) provide sufficient computing power for many applications in scientific as well as embedded applications. Unfortunately, when real-time, reliability, and security requirements need to be guaranteed, applications suffer from the interference with other applications, uncertainty of dynamic workload and state of the hardware. Composable application/architecture design and timing analysis is therefore a must for guaranteeing applications to satisfy their non-functional requirements independent from dynamic workload.

Invasive Computing can be used as the key enabler, as it provides the required isolation of resources allocated to each application. On the basis of this paradigm, this work presents a hybrid application mapping methodology that combines design-time analysis of application mappings with run-time management. Design space exploration delivers several resource reservation configurations with verified and/or validated non-functional properties of individual applications. These properties can then be guaranteed at run-time, as long as dynamic resource allocations comply with the offline analyzed resource configurations. In this work, we show that the approach provides increased flexibility and dynamism of systems even in the presence of hard real-time constraints. We also show the overhead of performing this dynamic application isolation based on run-time resource allocation.

5.20 Timing Verification – Flogging a Dead Horse?

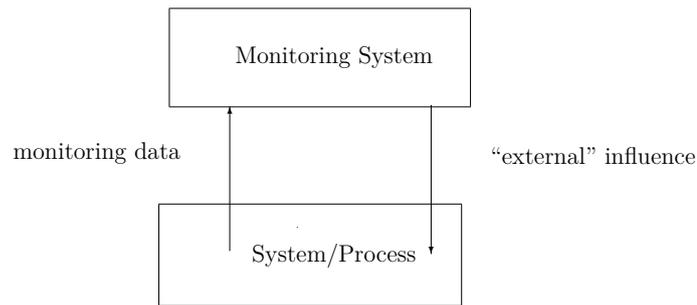
Reinhard Wilhelm (Universität des Saarlandes, DE)

License © Creative Commons BY 3.0 Unported license
© Reinhard Wilhelm

Joint work of Jan Reineke, Sebastian Hahn, Reinhard Wilhelm
Main reference P. Axer, R. Ernst, H. Falk, A. Girault, D. Grund, N. Guan, B. Jonsson, P. Marwedel, J. Reineke, C. Rochange, M. Sebastian, R. von Hanxleden, R. Wilhelm, W. Yi, “Building timing predictable embedded systems”, ACM Trans. Embed. Comput. Syst., Vol. 13(4), pp. 82:1–82:37, ACM, 2014.

URL <http://dx.doi.org/10.1145/2560033>

Recent developments in architectures and their adoption for safety- and time-critical embedded systems have reduced or even eliminated the chance to apply sound timing-verification techniques. The complexity, i.e., the size of the state space to be explored, has become just too large. The deployed new architectures have even made measurement-based, i.e. unsound techniques unpractical. What remains as alternative? The participants of this Dagstuhl Seminar have found no answer to this question. Maybe the timing-verification community should provide the proof that high-performance, predictable architectures are possible.



■ **Figure 1** Conceptual model of a runtime monitoring system.

6 Working groups

6.1 Runtime Monitoring

Felix Freiling (Universität Erlangen-Nürnberg, DE)

License Creative Commons BY 3.0 Unported license
© Felix Freiling

This breakout session ran for two days and focussed on the general topic of runtime monitoring for predictability and security. The participants on Wednesday were Lothar Thiele, Pieter Maene, Johannes Götzfried, Takeshi Sugawara, Jürgen Teich and Felix Freiling. On Thursday, the participants were Jan Reineke, Pieter Maene, Johannes Götzfried, Takeshi Sugawara, Jürgen Teich, Felix Freiling and Sudipta Chattopadhyay.

6.1.1 Conceptual model

We started with a conceptual model to clarify what we were talking about (see Figure 1). We usually have two systems: On the one hand there is the monitored system, which is the system whose properties we wish to monitor. If the monitored system would constantly satisfy its desired properties, there would be no necessity to monitor it. So we have the second system: the monitoring system. The monitoring system collects data of the monitored system (either digital or physical, e.g. using a camera) and has some internal logic that performs computation on this data. Under certain conditions, the monitoring system might influence the monitored system, e.g. by resetting the system or initiating a reconfiguration.

Formally, the observations of the monitoring system can be viewed as a sequence of events or states observed from the monitored system. Such sequences are often called *traces*. In a security context, the trustworthiness of the observed data is an issue. But also in a timeliness context the precision of the observed timings is critical to make truthful/good decisions. The classical examples of runtime monitoring systems are:

- watchdog timers,
- fault-injection attack countermeasures using sensors [2], or
- intrusion detection services within networks.

6.1.2 Issues in runtime monitoring

The discussion identified several issues that we used to structure the area of runtime monitoring systems. We discuss each of them separately in the following sections.

6.1.2.1 Types of properties that are monitored

We collected a set of properties that are on the usual wishlist of runtime monitoring systems:

- timeliness (i.e. meeting realtime deadlines),
- control flow integrity (i.e. the control flow does not deviate from the “correct” control flow intended by the programmer), or
- variance of response times (i.e., the variance of response times within the last hour is below a certain value).

It was noted that in general, properties that can be detected on individual traces fall in the class of safety properties (in the safety/liveness sense of Lamport [6] and Alpern and Schneider [1]). It is theoretically possible to derive bad events (or event sequences) from safety properties and to configure detection conditions for them.

It is well-known that certain types of functional and non-functional properties do not fall into the class of safety properties. For example, liveness properties (e.g., eventual termination) cannot be detected at runtime since they are only satisfied in infinite time (to detect violations one would have to observe the system infinitely long). But also variance of response times or information flow cannot be modeled as trace sets, but are a property of all traces of a system (see McLean [7]). They therefore cannot be checked at runtime. However, many such properties can be approximated by safety properties detectable at runtime (see work by Schneider [8]). For example, variance can be approximated using a finite time window in the past, or information flow can be approximated by also using the amount of entropy of the state trace of certain resources in the past.

Sometimes, it may be necessary or feasible to not detect precise events but rather observe “anomalies” of some form. This is the usual approach in intrusion detection where it is sometimes not clear how the attacker will attack and leave traces. If it is not totally clear what to detect, there are additional problems. For example, an attacker could try to avoid detection by trying to hide “beneath the radar”, e.g., by performing changes slowly enough so that they do not cause an anomaly.

6.1.2.2 Monitoring approaches

There can be different types of monitoring approaches. Which one is selected depends on the type of property being enforced and the type of fault/adversary/attacker being considered.

There is the distinction between permanent and periodic monitoring. Monitoring can also be done on demand or periodically but not in fixed intervals but in random (unpredictable) intervals. Given an intelligent adversary, periodic monitoring can be avoided because the attacker can wait for a check, then perform the attack and cover all traces before the next check occurs (sometimes called “ABA problem”).

In the literature, there is also the notion of a passive monitoring approach known as canaries. The idea is to detect fault using an artifact that is sensitive and easily broken by the fault:

- software canary (also known as cookie) which is used for buffer overflow protection like in StackGuard [4],
- active shield used to detect invasive probing of a circuit [3], and
- a special data store used to detect timing violation (i.e., canary flip flop).

6.1.2.3 Interactions between monitoring and monitored system

Usually it is assumed that monitoring and monitored system are subject to different types of faults/attacks. In security, this is realized through privilege separation (user/system level,

processor rings, ARM Trustzone, Intel SGX, etc.), such that the attacker assumption can be justified that the privileged area is not affected by the attack. Without such a restriction on the attacker no security can be guaranteed (see breakout discussions on “attacker models”).

In the safety/realtime/fault-tolerance area it is sometimes the case that two systems take on the role of monitoring and monitored system for each other (see for example self-checking checkers from coding theory or fault-tolerant protocols for voting or agreement). In such scenarios the fault assumption is restricted in the way that both monitoring and monitored system can be affected by faults but not both at the same time. In this context, we often find fault/attacker assumptions such as “ k -out-of- n ” (meaning that at most k out of n systems are affected by faults).

6.1.2.4 Suitable reactions

In case the monitoring system issues a reaction, what can this be? In fault-tolerant systems there exists the notion of fail-safe meaning that the system is switched from operational to a safe state (e.g. “all signals stop” in railway systems). This is problematic in situations where there is no safe alternative to normal operation (as in avionics). In this case, the minimum you can do is to at least ensure that sufficient evidence is collected for a later (post mortem) analysis, i.e. work which has been done under the heading of secure logging.

In security there is the notion of fail-secure meaning that if the system fails, security properties like integrity or confidentiality are not violated or re-established (e.g., forward secrecy). Interestingly, availability is not very often investigated in a security context. Availability, however, is important in this context since reactions to faults/attacks usually mean a form of reset or restart, and continuous triggers to restart can cause unavailability (denial-of-service attacks). This is especially problematic when unavailability is the result of false detections (false positives) of the monitoring system.

A suitable reaction usually is to adjust resources necessary for the task and continue the task with better resources. Continuation can mean that it restarts from a previous (uncorrupted) checkpoint. In this context, the notion of stabilization was mentioned [5] meaning that as long as faults/attacks happen, no progress is guaranteed, but that the system automatically regains progress once faults/attacks stop to occur.

6.1.3 Open problems

While most of the above observations can be considered known or at least named in the literature, we collected a couple of open points that were considered novel aspects of the problem that deserved some research attention:

- The whole issue of *distributed monitoring* has its problems of its own: Distribution creates challenges with the attacker model, trust issues in the exchange of information, problems of global observation etc.
- If a monitoring system is needed, why not generate it automatically? Given a monitoring property, can we automatically generate a monitoring system in software and/or hardware that observes it?
- What is the correct/right granularity of checking the monitored property? Should monitoring be done periodically, on demand or continuously? What is the relation to efficiency of the monitoring process?
- Monitoring usually assumes that there is some information against which a monitored property can be checked? Can this signature be normalized? Can it possibly be reduced to the knowledge of a (cryptographic) key?

- Adding monitoring functionality increases the attack surface of the program. To what extent does the monitoring functionality affect security then?
- In what sense are safety properties also security properties? Obviously this depends on the attacker model: Which attacker model comprises which fault model? In case an attacker model includes a fault model, in what way does runtime monitoring for faults subsume monitoring effort for attacks? These questions refer to the synergies of faults vs. attack detection.

References

- 1 B. Alpern and F. B. Schneider. Defining liveness. *Information Processing Letters*, 21:181–185, 1985.
- 2 Josep Balasch. Introduction to fault attacks. Presentation at IACR Summer School, Chia Laguna, Sardinia, October 2015. https://www.cosic.esat.kuleuven.be/summer_school_sardinia_2015/slides/Balasch.pdf.
- 3 Sébastien Briaïs, Stéphane Caron, Jean-Michel Cioranescu, Jean-Luc Danger, Sylvain Guilley, Jacques-Henri Jourdan, Arthur Milchior, David Naccache, and Thibault Porteboeuf. 3D hardware canaries. In Emmanuel Prouff and Patrick Schaumont, editors, *Cryptographic Hardware and Embedded Systems – CHES 2012 – 14th International Workshop, Leuven, Belgium, September 9-12, 2012. Proceedings*, volume 7428 of *Lecture Notes in Computer Science*, pages 1–22. Springer, 2012.
- 4 Crispan Cowan. Stackguard: Automatic adaptive detection and prevention of buffer-overflow attacks. In Aviel D. Rubin, editor, *Proceedings of the 7th USENIX Security Symposium, San Antonio, TX, USA, January 26-29, 1998*. USENIX Association, 1998.
- 5 E. W. Dijkstra. Self-stabilizing systems in spite of distributed control. *Communications of the ACM*, 17(11):643–644, November 1974.
- 6 Leslie Lamport. Proving the correctness of multiprocess programs. *IEEE Transactions on Software Engineering*, 3(2):125–143, March 1977.
- 7 John McLean. A general theory of composition for a class of “possibilistic” properties. *IEEE Transactions on Software Engineering*, 22(1):53–67, January 1996.
- 8 Fred B. Schneider. Enforceable security policies. *ACM Transactions on Information and System Security*, 3(1):30–50, February 2000.

6.2 Future of Timing Verification

Samarjit Chakraborty (TU München, DE)

License © Creative Commons BY 3.0 Unported license
© Samarjit Chakraborty

The breakout session explored the future of timing verification of real-time systems. This breakout session was motivated by the fact safety-critical embedded systems are increasingly relying on advanced processor architectures that have been designed with the goal of improving average-case performance and not predictability. Hence, while there have been considerable advancements in the domains of Worst Case Execution Time (WCET) analysis of programs, and also timing analysis of real-time systems, this is increasingly appearing to be a losing battle. More importantly, academic research in the domains of WCET and timing analysis has had little impact on practice. Therefore, the question is what is the path forward from here?

Here, one line of thought that has emerged during the last 1-2 years, especially in the context of embedded control systems is how many deadlines and which ones should really

be met in a real-time system? In other words, is 100% predictability, as aimed in real-time systems research really needed? Typically, many safety critical systems implement some control algorithm. Meeting control performance requirements are subject to satisfying some timing constraints, which a real-time systems theorist tries to verify (schedulability analysis) or ensure (schedule design or synthesis). Hence, deadline constraints have served as a good interface between control theorists and real-time/embedded systems theorists.

However, most feedback control systems have an inherent degree of robustness. Also when the plant is in a “steady” state, the system can be run in open loop (i.e., no computation of control input is necessary). This means that even if some control signals are not computed or transmitted in time, the control performance still remains acceptable. If these control signals can be characterized, i.e., acceptable patterns of deadline misses may be computed then they might be interpreted as a quantification of the degree of predictability that is needed.

This means that timing analysis, instead of focusing on deadline constraints, should focus on higher-level (control theoretic) goals that better characterize the systems performance requirements. However, computing acceptable patterns of deadline violations is not trivial and requires sophisticated control theoretic analysis. For example, see [1, 2]. Further, timing or schedulability analysis to ascertain that at most these deadline violations may happen is more difficult than checking that no deadline violations happen. Nevertheless, such an approach gives a certain leeway that could be exploited to allow platforms and architectures that cannot be completely analyzed but instead only certain timing bounds on their performance may be given.

The discussion during this breakout session was also on the need for end-to-end timing analysis, e.g., considering the influence of operating systems on the execution time of code, which has not been sufficiently addressed until now. Finally, the need for benchmarks and the reproducibility of timing analysis techniques were also discussed. One potential solution would be to consider Simulink models of different controllers (e.g., from the automotive domain) and use the code synthesized from these models for timing analysis.

References

- 1 Dip Goswami, Reinhard Schneider, Samarjit Chakraborty. Relaxing Signal Delay Constraints in Distributed Embedded Controllers. *IEEE Trans. Control Systems Technology* 22(6): 2337-2345, 2014
- 2 Dip Goswami, Samarjit Chakraborty, Purandar Bhaduri, Sanjoy K. Mitter. Characterizing feedback signal drop patterns in formal verification of networked control systems. *IEEE International Symposium on Computer-Aided Control System Design (CACSD)*, 2013

6.3 Attack Models

Albert Cohen (ENS – Paris, FR) and Karine Heydemann (UPMC – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Albert Cohen and Karine Heydemann

This breakout session focused on a general survey of attack models. The participants were Albert Cohen, Ruan De Clercq, Felix Freiling, Gernot Heiser, Karine Heydemann, Patrick Koeberl, Peter Maene, Claire Maiza, Sibin Mohan, Frank Mueller, Patrick Schaumont, and Takeshi Sugawara.

System designers need to define security properties and protective measures to implement them. This process involves systematic characterization of attack models. The working group

conducted survey of logical and physical attacks targeting CPS and IoT devices. The focus was on threats and attack models in general, in those associated with multi- and many-core systems in particular. Building on such a survey, researchers will be able to determine and to quantify the aspects of the design and implementation methods that need to be revisited, to integrate the security dimension at the heart of defense in depth mechanisms, correct-by-construction design, test, verification, validation, and certification.

6.4 Synergy between Predictability and Security

Frank Mueller (North Carolina State University – Raleigh, US)

License  Creative Commons BY 3.0 Unported license
© Frank Mueller

This break-out session was attended by about 25 participants and included other break-out ideas on availability, multi-cores, and levels on security.

In an initial brain-storming session, the full breadth of the topic was covered, where each participant contributed their ideas. The second session was dedicated to intensive discussions on various topics and concluded by various action items, including a security attack contest using predictability as a means of attack. The brain-storming ideas and following discussions are summarized under a number of topic headings:

6.4.1 Predictability versus Security

Participants observed that some timing predictability techniques are synergistic with security (and vice versa) while others are antagonistic in the sense that they appear to be in direct conflict. Synergistic examples range from timing information already available due to real-time analysis used for intrusion detection over obfuscation techniques in scheduling to the simplicity of crypto-algorithms facilitating their timing predictability. Antagonistic examples range from real-time predictability that may facilitate attacks over timing faults as a means to attack systems to a discussion on whether or not randomized attacks completely void any attempts to increase software diversity (including but not limited to parallelism/scheduling) when the number of variants is fixed.

6.4.2 Parallelism

One discussion was dedicated to the challenges of parallelism with diverse opinions. While isolation (in space, e.g., via partitioning, or in time, e.g., via TDMA) helps both predictability and security most of the time, it may come at a certain cost. For example, MMU protection results in page faults, which are costly, yet certain data structures may not require stringent access protection via MMUs (or MCUs), i.e., the idea of different levels of security seems intriguing.

6.4.3 Side-channel Attacks

The discussion centered around software-based side-channel attacks. Different types of leaked information (especially with respect to timing) and counter-measures were discussed. Mitigation techniques, e.g., obfuscation via randomization, were noted to adversely affect predictability while isolation typically aids predictability.

6.4.4 Availability

It was noted that no viable solution for predictability appears to exist. Once compromised, a hardware unit can always be powered down, even without physical access (assuming firmware safe-guards can be circumvented or are also affected by the intrusion). The idea of an analogy to priority inversion in security of a lower critical task affecting a higher critical one was discussed. Containment methods to handle faults by switching to more simplistic, higher protection modes also seem attractive in this context. But more fundamental work is required to better understand this subject.

6.4.5 Connectivity

Not much time was spent on discussing connectivity modes (on/off), but it was noted that isolation in proprietary networks can help. Independently, a need for protecting edge devices (e.g., IoT) was voiced as they, in large numbers, can easily orchestrate a DDOS attack.

6.4.6 Wish List

We formulated a wish list of action items. One challenge is to come up with a 3-dimensional model that combines security, predictability and also safety/fault tolerance as all three are inter-connected. Another challenge is the need for a hierarchy/levels, possibly for each of these three areas, and/or in a combined model. Different degrees of protection, degrees of information leakage, affected software mechanisms etc. may require a different response in protection. But most of all, security should be a first-order design principle, not an after-thought, as it currently is. And while fundamental security research is still required that may lead to completely new design methods, research is nonetheless needed to devise methods for retrofitting security into existing hardware/software systems.

6.4.7 The Programming Challenge

A final discussion culminated in the idea of posing a programming challenge for a successful timing attack, possibly in an Autosar setting. If successful, extensive PR should be used to put timing problems into the limelight within the realm of security, much in line with the way that the real-time system community received credit for saving the Mars Lander mission due to priority inheritance support on its software platform. One open question was if a follow-on 1-week attack hackathon should be organized to achieve this goal.

6.5 Models of Computation and Programming Languages

Reinhard von Hanxleden (Universität Kiel, DE)

License  Creative Commons BY 3.0 Unported license
© Reinhard von Hanxleden

There exist a multitude of models of computation (MoCs) and associated programming languages. Clearly, there is not a single winner, each has its pros and cons, and often a combination of languages is used in different parts of the design and at different abstraction levels. The aim of this break out session, based on two proposals brought in by David Broman and Reinhard von Hanxleden, was to identify the specific issues that arise for adaptive isolation. A number of questions were identified in advance, such as which the

essential language constructs are, which concepts should be left out, and what can we learn from currently available real-time languages and APIs. Given the setting of the seminar on multiprocessor systems on chip (MPSoCs), the role of concurrency was also brought forward as a possible focus. The participants in the breakout session were Davide Bertozzi (Università di Ferrara, IT), David Broman (KTH Stockholm, SE), Reinhard von Hanxleden (U Kiel, DE), Frank Mueller (North Carolina State University – Raleigh, US), Zoran Salcic (University of Auckland, NZ), Martin Schoeberl (Technical University of Denmark – Lyngby, DK), Stefan Wildermann (Universität Erlangen-Nürnberg, DE) and the aforementioned proposers.

As it turned out, the session participants were mostly from the predictability field, thus the resulting discussions centered around that aspect and had little focus on security. Furthermore, rather than trying to propose specific MoCs and languages, the participants agreed to try to identify properties that suitable MoCs and languages should have (even though specific MoCs/languages were covered to some extent).

To start with, possible application areas were discussed, including industrial control, image processing such as in a lane following assistant, and generally cyber-physical systems. There was a question mark on whether also financial applications would be within the scope, but later it was agreed that these would possess similar qualities.

Next, it was discussed what we try to avoid. There are obviously non-desirable scenarios, such as catastrophic failures in power plants and the like, but also “blue screens” that indicate undesired system states and loss of functionality and consequently may lead to a bad product reputation and resulting economic consequences. On a more technical level, a MoC/language should try to rule out things like memory leaks, timing issues due to garbage collection, and vulnerabilities such as buffer overflows. One participant also brought forward that “C programming” should be avoided; this, taken literally, would be a rather difficult proposition, given that C is still one of the most popular languages in particular in the aforementioned application domains. However, it was agreed (and that was the point of that proposal) that some of the programming constructs offered by C and similar languages are to be used with care.

The proper usage of languages like C led to the next topic, namely what we try to aim for. An MoC/language should be intuitive and familiar. However, in particular for languages that were not specifically designed with the above listed goals in mind, these languages should be used in a disciplined manner, possibly using subsets such as e.g. MISRA-C proposed by the Motor Industry Software Reliability Association. For example, ruling out dynamic memory allocation could be used to prevent memory leaks. Furthermore, one might want to use C etc. merely as an intermediate language, to be synthesized from a higher-level modeling language. The semantics should also be agnostic to available resources, as far as possible; e.g., a program should behave the same (apart from performance issues) regardless of on how many cores it is executed. Another important aspect for a language is to make a clear separation between the end user, the low-level machine implementation expert, and the compiler writer. Such separation of concern may make it possible for end-users to develop efficient systems (performance and predictable) within a shorter development time frame.

As obstacles for achieving the above goals the participants identified for example processors with unpredictable timing, as was also emphasized during earlier presentations (e.g. by Reinhard Wilhelm) during the seminar. However, it was agreed upon that the whole design flow matters, not only the execution platform. For example, a programming language should have full control over reactive control flow (concurrency and preemption), instead of handing over that responsibility to the operating system or some run time system such as the Java Virtual Machine. Typically, application-level control tasks do have a certain

level of robustness; e.g., an airplane usually can tolerate if a control output arrives after 6 msec instead of 5 msec or is missing altogether for a cycle or two. Similarly, an extra cache miss or pipeline stall should not lead to a significant change in system behavior. Design and synthesis paths must be robust, not brittle, and the used MoCs/languages should provide for that.

To conclude, some questions and remaining challenges were posed. How should time be incorporated? Do we need new languages, new MoCs? Do DSLs help? Should we have more open OSs that take hints from user/compiler? How do we achieve composability, e.g. preserve functional as well as non-functional properties of individual components?

7 Panel Discussion

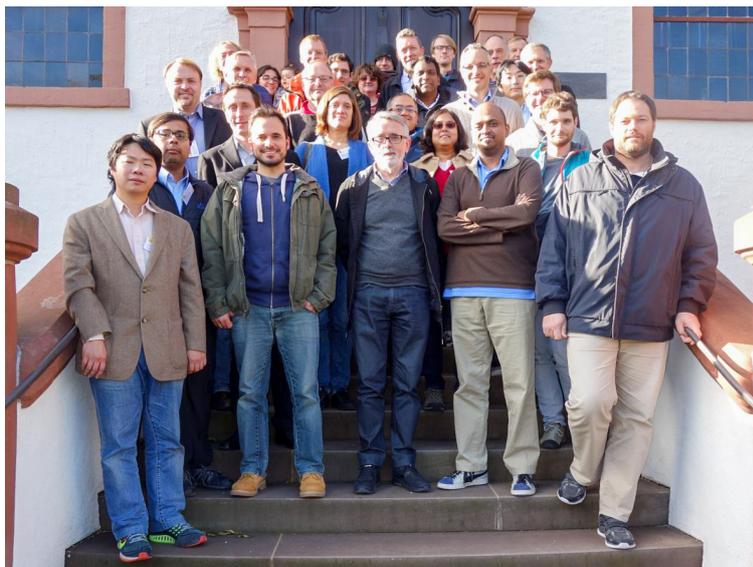
Sri Parameswaran (UNSW) organized a panel discussion on adaptive isolation. The panelists were Patrick Koeberl (Intel), Tulika Mitra (NUS), Jürgen Teich (Friedrich-Alexander-Universität Erlangen-Nürnberg), and Lothar Thiele (ETH Zürich). The discussion centered around motivation for the seminar, the similarities, differences between security and predictability and their interactions/impact in the context of isolation, the future of timing predictability in the absence of commercial predictable architectures, and the tools, techniques, mechanisms for isolation covered in the seminar.

8 Acknowledgements

We would like to take the opportunity to thank and acknowledge our organization team member Ingrid Verbauwhede for her great effort in contributing brilliant ideas and suggestions on the topic of the seminar as well as to the list of participants. For very unexpected reasons, she could unfortunately not participate in the seminar.

Participants

- Davide Bertozzi
Università di Ferrara, IT
- Björn B. Brandenburg
MPI-SWS – Kaiserslautern, DE
- David Broman
KTH Royal Institute of
Technology, SE
- Samarjit Chakraborty
TU München, DE
- Sudipta Chattopadhyay
Universität des Saarlandes, DE
- Jian-Jia Chen
TU Dortmund, DE
- Albert Cohen
ENS – Paris, FR
- Ruan de Clercq
KU Leuven, BE
- Heiko Falk
TU Hamburg-Harburg, DE
- Felix Freiling
Univ. Erlangen-Nürnberg, DE
- Johannes Götzfried
Univ. Erlangen-Nürnberg, DE
- Gernot Heiser
UNSW – Sydney, AU
- Andreas Herkersdorf
TU München, DE
- Karine Heydemann
UPMC – Paris, FR
- Patrick Koeberl
Intel – Hillsboro, US
- Pieter Maene
KU Leuven, BE
- Claire Maiza
Université Grenoble Alpes –
Sait Martin d’Hères, FR
- Peter Marwedel
TU Dortmund, DE
- Tulika Mitra
National University of
Singapore, SG
- Sibin Mohan
Univ. of Illinois – Urbana, US
- Frank Mueller
North Carolina State University –
Raleigh, US
- Sri Parameswaran
UNSW – Sydney, AU
- Jan Reineke
Universität des Saarlandes, DE
- Christine Rochange
University Toulouse, FR
- Zoran Salcic
University of Auckland, NZ
- Patrick Schaumont
Virginia Polytechnic Institute –
Blacksburg, US
- Martin Schoeberl
Technical University of Denmark
– Lyngby, DK
- Wolfgang Schröder-Preikschat
Univ. Erlangen-Nürnberg, DE
- Takeshi Sugawara
Mitsubishi – Kanagawa, JP
- Jürgen Teich
Univ. Erlangen-Nürnberg, DE
- Lothar Thiele
ETH Zürich, CH
- Theo Ungerer
Universität Augsburg, DE
- Reinhard von Hanxleden
Universität Kiel, DE
- Stefan Wildermann
Univ. Erlangen-Nürnberg, DE
- Reinhard Wilhelm
Universität des Saarlandes, DE



Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR)

Edited by

Roger K. Moore¹, Serge Thill², and Ricard Marxer³

1 University of Sheffield, GB, r.k.moore@sheffield.ac.uk

2 University of Skövde, SE, serge.thill@his.se

3 University of Sheffield, GB, r.marxer@sheffield.ac.uk

Abstract

This seminar was held in late 2016 and brought together, for the first time, researchers studying vocal interaction in a variety of different domains covering communications between all possible combinations of humans, animals, and robots. While each of these sub-domains has extensive histories of research progress, there is much potential for cross-fertilisation that currently remains underexplored. This seminar aimed at bridging this gap. In this report, we present the nascent research field of VIHAR and the major outputs from our seminar in the form of prioritised open research questions, abstracts from stimulus talks given by prominent researchers in their respective fields, and open problem statements by all participants.

Seminar October 30–4, 2016 – <http://www.dagstuhl.de/16442>

1998 ACM Subject Classification I.2.7 Natural Language Processing, I.2.9 Robotics

Keywords and phrases animal calls, human-robot interaction, language evolution, language universals, speech technology, spoken language, vocal expression, vocal interaction, vocal learning

Digital Object Identifier 10.4230/DagRep.6.10.154

1 Executive Summary

Serge Thill

Ricard Marxer

Roger K. Moore

License  Creative Commons BY 3.0 Unported license
© Serge Thill, Ricard Marxer, and Roger K. Moore

Almost all animals exploit vocal signals for a range of ecologically-motivated purposes. For example, predators may use vocal cues to detect their prey (and vice versa), and a variety of animals (such as birds, frogs, dogs, wolves, foxes, jackals, coyotes, etc.) use vocalisation to mark or defend their territory. Social animals (including human beings) also use vocalisation to express emotions, to establish social relations and to share information, and humans beings have extended this behaviour to a very high level of sophistication through the evolution of speech and language – a phenomenon that appears to be unique in the animal kingdom, but which shares many characteristics with the communication systems of other animals.

Also, recent years have seen important developments in a range of technologies relating to vocalisation. For example, systems have been created to analyse and playback animals calls, to investigate how vocal signalling might evolve in communicative agents, and to interact with users of spoken language technology (voice-based human-computer interaction using speech technologies such as automatic speech recognition and text-to-speech synthesis). Indeed, the



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

Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR), *Dagstuhl Reports*, Vol. 6, Issue 10, pp. 154–194

Editors: Roger K. Moore, Serge Thill, and Ricard Marxer



DAGSTUHL
REPORTS Dagstuhl Reports

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

latter has witnessed huge commercial success in the past 10-20 years, particularly since the release of *Naturally Speaking* (Dragon's continuous speech dictation software for a PC) in 1997 and Siri (Apple's voice-operated personal assistant and knowledge navigator for the iPhone) in 2011. Research interest in this area is now beginning to focus on voice-enabling autonomous social agents (such as robots).

Therefore, whether it is a bird raising an alarm, a whale calling to potential partners, a dog responding to human commands, a parent reading a story with a child, or a businessperson accessing stock prices using an automated voice service on their mobile phone, vocalisation provides a valuable communications channel through which behaviour may be coordinated and controlled, and information may be distributed and acquired.

Indeed, the ubiquity of vocal interaction has given rise to a wealth of research across an extremely diverse array of fields from the behavioural and language sciences to engineering, technology and robotics. This means that there is huge potential for crossfertilisation between the different disciplines involved in the study and exploitation of vocal interactivity. For example, it might be possible to use contemporary advances in machine learning to analyse animal activity in different habitats, or to use robots to investigate contemporary theories of language grounding. Likewise, an understanding of animal vocal behaviour might inform how vocal expressivity might be integrated into the next generation of autonomous social agents. Some of these issues have already been addressed by relevant sub-sections of the research community. However, many opportunities remain unexplored, not least due to the lack of a suitable forum to bring the relevant people together.

Our Dagstuhl seminar on the topic of "Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR)" provided the unique and timely opportunity to bring together scientists and engineers from a number of different fields to appraise our current level of knowledge. Our broad aim was to focus discussion on the general principles of vocal interactivity as well as evaluating the state-of-the-art in our understanding of vocal interaction within-and-between humans, animals and robots. Some of these sub-topics, such as human spoken language or vocal interactivity between animals, have a long history of scientific research. Others, such as vocal interaction between robots or between robots and animals, are less well studied – mainly due to the relatively recent appearance of the relevant technology. What is interesting is that, independent of whether the sub-topics are well established fields or relatively new research domains, there is an abundance of open research questions which may benefit from a comparative interdisciplinary analysis of the type addressed in this seminar.

2 Table of Contents

Executive Summary

| | |
|---|-----|
| <i>Serge Thill, Ricard Marxer, and Roger K. Moore</i> | 154 |
|---|-----|

Seminar Organization

| | |
|---|-----|
| Participants | 158 |
| Overall organisation | 158 |
| Prioritisation of open research questions | 158 |
| Conclusions and next steps | 160 |

Overview of Talks

| | |
|--|-----|
| (Vocal) interaction with the artificial <i>Tony Belpaeme</i> | 160 |
| Acoustic communication in animals: a window to their inner state <i>Elodie Briefer</i> | 160 |
| Lessons about Vocal Interaction from Joint Speech, or How I learned to love linguaging <i>Fred Cummins</i> | 162 |
| Universals in Vertebrate Vocal Communication <i>Angela Dassow</i> | 162 |
| Temporal models for machine listening in mixed audio scenes <i>Dan Stowell</i> | 163 |

Open Problems

| | |
|--|-----|
| Statement by Andrey Anikin <i>Andrey Anikin</i> | 164 |
| Statement by Timo Baumann <i>Timo Baumann</i> | 166 |
| Statement by Tony Belpaeme <i>Tony Belpaeme</i> | 167 |
| Statement by Elodie Briefer <i>Elodie Briefer</i> | 168 |
| Statement by Nick Campbell <i>Nick Campbell</i> | 169 |
| Statement by Fred Cummins <i>Fred Cummins</i> | 171 |
| Statement by Angela Dassow <i>Angela Dassow</i> | 172 |
| Statement by Robert Eklund <i>Robert Eklund</i> | 173 |
| Statement by Julie E. Elie <i>Julie E. Elie</i> | 176 |

Statement by Sabrina Engesser
Sabrina Engesser 177

Statement by Sarah Hawkins
Sarah Hawkins 179

Statement by Ricard Marxer
Ricard Marxer 181

Statement by Roger K. Moore
Roger K. Moore 182

Statement by Julie Oswald
Julie Oswald 183

Statement by Bhiksha Raj
Bhiksha Raj 184

Statement by Rita Singh
Rita Singh 185

Statement by Dan Stowell
Dan Stowell 186

Statement by Zheng-Hua Tan
Zheng-Hua Tan 188

Statement by Serge Thill
Serge Thill 189

Statement by Petra Wagner
Petra Wagner 190

Statement by Benjamin Weiss
Benjamin Weiss 192

Participants 194

3 Seminar Organization

3.1 Participants

Participants at the seminar spanned the entire range of academic career stages and provided broad global coverage. The mix of attendants was also particularly interdisciplinary, covering animal vocalisation, human language, machine language production and understanding, as well as various intersections of these topics. The success of the seminar can largely be attributed to this presence of multi- and pluridisciplinary interests and the discussions across boundaries that this generated

3.2 Overall organisation

We intentionally kept the structure of the seminar rather loose and self-organising. We started the seminar with brief presentations from all participants in which they were asked to summarise their background, and what they felt the most pressing issues were. This was then followed by invited “stimulus” talks by selected participants from different home disciplines. The intention with these talks was to foster an initial interdisciplinary interest between the different communities present.

The remainder of the seminar was spent in groups that focussed on specific issues thus identified, which spanned a broad range of topics, as can be seen in the remainder of this report.

3.3 Prioritisation of open research questions

The seminar was guided by a recent review paper by the organisers, which identified a number of open research questions in the field of VIHAR. During the seminar, we asked the participants to identify the three questions they considered most crucial / relevant. This allows us to here present the resulting order of questions that received at least one vote:

1. *(16 votes)*
 - What is the relationship (if any) between language and the different signalling systems employed by non-human animals?
2. *(11 votes)*
 - What tools might be needed in the future to study vocalisation in the wild?
3. *(10 votes)*
 - What are the similarities/differences between the vocal systems (including brain organisation) in different animals?
4. *(8 votes)*
 - How does one evolve the complexity of voice-based interfaces from simple structured dialogues to more flexible conversational designs without confusing the user?
 - Are there any mathematical modelling principles that may be applied to all forms of vocal interactivity and is it possible to derive a common architecture or framework for describing vocal interactivity?
5. *(7 votes)*
 - What are the limitations (if any) of vocal interaction between non-conspecifics?
6. *(6 votes)*
 - How can vocal interactivity as an emergent phenomenon be modelled computationally?

7. (5 votes)
 - What are the common features of vocal learning that species capable of it share, and why is it restricted to only a few species?
8. (4 votes)
 - What are the common features of vocal learning that species capable of it share, and why is it restricted to only a few species?
 - How are vocal mechanisms constrained or facilitated by the morphology of the individual agents involved?
 - How is information distributed across the different modes and what is the relationship between vocal and non-vocal (sign) language?
 - To what degree can affective states be interpreted and expressed, and should they be treated as superficial or more deeply rooted aspects of behaviour?
 - Do the characteristics of vocalisations carry information about the social relationship connecting the interactants (for example, how is group membership or social status signalled vocally)?
9. (3 votes)
 - Do artificial agents need ToM in order to interact effectively with human beings vocally?
 - How are multi-modal behaviours orchestrated, especially in multi-agent situations?
 - Who should adapt to whom in order to establish an effective channel?
 - How would one model the relevant dynamics (whether to study natural interactivity or to facilitate human-machine interaction)?
 - How can insights from such questions inform the design of vocally interactive artificial agents beyond Siri?
10. (2 votes)
 - How are vocalisations manipulated to achieve the desired results and is such behaviour reactive or proactive?
 - Is ToM crucial for language-based interaction?
 - To what degree is there a phonemic structure to animal communications, and how would one experimentally measure the complexity of vocal interactions (beyond information-theoretic analyses)?
 - What is it about the human-dog relationship that makes the one-sidedness of this relation sufficient, and conversely, what can biases in communication balancing say about social relationships?
 - To what extent are vocal signals teleological, and is it possible to distinguish between intentional and unintentional vocalisations?
 - Given the crucial nature of synchrony and timing in interactivity between natural agents, to what extent does this importance carry over to human-machine dialogue?
 - Is it necessary to create new standards in order to facilitate more efficient sharing of research resources?
11. (1 vote)
 - What is the role of vocal affect in coordinating cooperative or competitive behaviour?
 - How does a young animal (such as a human child) solve the correspondence problem between the vocalisations that they hear and the sounds that they can produce?
 - Does the existence (or absence) of prior relationships between agents impact on subsequent vocal activity?
 - How is vocalisation used to sustain long-term social relations?
 - What can be learned from attempts to teach animals human language (and vice versa)?

3.4 Conclusions and next steps

Overall, the seminar proved very successful in the fostering of a new community targetting the interdisciplinary challenges in the field of VIHAR. we now intend to keep the momentum going and, as a next step, are organising a workshop on VIHAR as a satellite event of Interspeech 2017 (see <http://vihar-2017.vihar.org/> for details). This workshop is a direct consequence of the Dagstuhl seminar; indeed its organising committee consists of participants in the seminar. In conclusion, we hope that, with this seminar, we have laid the foundations for a new and vibrant research community that will remain active and meet regularly for years to come.

4 Overview of Talks

4.1 (Vocal) interaction with the artificial

Tony Belpaeme (University of Plymouth, GB)

License  Creative Commons BY 3.0 Unported license
© Tony Belpaeme

Artificial Intelligent systems, from digital assistants to humanoid robots, are already part and parcel of our daily lives or are expected to be in the not too distant future. When interacting with artificial systems, we now use channels different to those we use to interact with others. We type commands and search terms, but do not yet engage in dynamic social interactions with machines. It is however interesting to see how anthropomorphisation colours our interactions with machines: we cannot help interpret behaviours of machines (even the simplest Braitenberg vehicles) as having intention and personality, and this can be exploited by engineers when building socially interactive robots. I will present two studies, the first study looks into how people interpret robotic clicks and beeps (such as the sounds made by R2D2). These non-linguistic utterances are readily interpreted as containing emotion, and adults interpret these sounds categorically. The second study shows how a robot can leverage the human propensity to tutor: using a social robot we study how people teach it, and show that they form a mental model of the robot used to tailor the teaching experiences for the robot. Both studies not only show how human social interaction spills over to interacting with machines, but also demonstrate the promise of using robots as research tools, achieving a level of control and repeatability unachievable by current research methods in human and animal interaction.

4.2 Acoustic communication in animals: a window to their inner state

Elodie Briefer (ETH Zürich, CH)

License  Creative Commons BY 3.0 Unported license
© Elodie Briefer

My presentation focuses on three main questions, relevant for VIHAR (Moore et al. 2016 Front. Robot. AI 3:61); 1) To what degree is there a phonemic structure to animal communications, and how would one experimentally measure the complexity of vocal interactions (beyond information-theoretic analyses)?; 2) To what degree can affective states be interpreted

and expressed, and should they be treated as superficial or more deeply rooted aspects of behavior?; 3) What are the limitations (if any) of vocal interaction between non-conspecifics?

1. To what degree is there a phonemic structure to animal communications, and how would one experimentally measure the complexity of vocal interactions (beyond information-theoretic analyses). I discuss the first question through my PhD research on skylarks. Skylark songs are among the most complex acoustic signals compared to other songbird species, both in terms of the number of different acoustic units produced by each male, but also in how they are arranged within songs (high diversity of transitions). Markov chain analyses show that skylark's song are best modelled by a 1st order Markov chain governed by a finite-state grammar. However, as each acoustic unit bears no "meaning" per se (different units do not have different referential meaning), this structure could thus be described as "phonetic patterning". In addition, geographical variation exists at the sequential level; sequences of syllables are shared by neighbouring birds. Playback experiments revealed that the dialect constitutes a group signature used by birds to discriminate neighbours (birds from the same group) from strangers (birds from a different group), and that the order of acoustic units within these particular sequences is an important feature for the neighbourhood identity coding.
2. To what degree can affective states be interpreted and expressed, and should they be treated as superficial or more deeply rooted aspects of behavior? I discuss the second question through my current research on vocal expression of emotions. Expression of emotions plays an important role in social species, including humans, because it regulates social interactions. Indicators of emotions in human voice have been studied in detail. However, similar studies testing a direct link between emotions and vocal parameters in non-human animals are rare. In particular, little is known about how animals encode in their vocalisations, information about the valence (positive/negative) of the emotion they are experiencing. I combined new frameworks recently adapted from humans to animals to analyse vocalisations (source-filter theory), and emotions (dimensional approach), in order to decipher vocal expression of both arousal (bodily activation) and valence in domestic ungulates. I present my results on horses, which are part of a large project aimed at investigating the evolution of vocal expression of emotions in ungulates (goats, horses, pigs and cattle) and the effect of domestication on human-animal communication. I measured physiological, behavioural and vocal responses of the animals to several situations characterised by different emotional arousal and valence. Physiological and behavioural measures collected during the tests confirmed the presence of different underlying emotions. My results showed that horse whinnies are composed of two fundamental frequencies (two "voices"), suggesting biphonation, a rare case among mammals. Interestingly, one of these fundamental frequency and the energy spectrum indicates emotional arousal, while the other and the duration indicates the emotional valence of the producer. These findings show that cues to emotional arousal and valence are segregated in different, relatively independent parameters of horse whinnies. Most of the emotion-related changes to vocalisations that I observed are similar to those observed in humans and other species, suggesting that vocal expression of emotions has been conserved throughout evolution.
3. What are the limitations (if any) of vocal interaction between non-conspecifics? To discuss the third question, I am showing current experiments that I am currently running, testing if domestic and wild horses perceive indicators of emotions in conspecific vocalisations, vocalisations of closely related species (wild horses to domestic horses and vice versa), as well as in human voice. I am also testing if humans can perceive emotions in the

vocalisations of domestic (goats, horses, pigs and cattle) and wild (Przewalski's horses and wild boars) ungulates using an online questionnaire. These experiments will shed light on the evolution of vocal expression of emotions, and on the impact of domestication on human-animal communication of emotions. From this work, more questions arise, and an inter-disciplinary approach joining research on vocal communication within and between animals, humans and robots would be greatly beneficial to share tools and skills, in order to lead to further advances in these fields of research.

4.3 Lessons about Vocal Interaction from Joint Speech, or How I learned to love languaging

Fred Cummins (University College Dublin, IE)

License  Creative Commons BY 3.0 Unported license
© Fred Cummins

The theme of this seminar encourages us to look beyond the well-studied terrain of inter-human communication towards some novel and unexplored challenges. This attitude seems to suggest that inter-human communication is sufficiently well-studied to support generalisation, but I will argue that there is plenty of reason to believe that we have failed to identify language as an object of study. The shortcomings of received approaches to language are graphically illustrated by the wholly Christian and literate foundation of orthodox accounts of language that ignore such essential parts of the fabric of communication as (1) gesture, (2) gaze, (3) posture, (4) prosody and my favourite topic, (5) joint speech. Joint speech is speech produced by multiple speakers all saying the same thing at the same time, as found in practices of prayer, ritual, protest, sports traditions, and beyond. The study of such common and important practices throws up some conundrums and sensitises us to some themes notably absent from most of contemporary linguistics: Common Ground emerges as an essential concept for understanding the dialogical push and pull of many interactions. This may be relevant when we come to consider human-animal communication, where it may find expression in the Bayesian language of shared priors. Co-presence is an important theme that gets lost when we approach languaging from a representationalist perspective. The importance of context, rather than lexis and text, jumps out at us, and in place of an exegesis based on the analysis of encoded messages, we are encouraged to look instead at the role of real-time reciprocal interaction. Thus joint speech may alert us to some blind spots we have in accounts of inter-human communication that will gain significance as our study extends to communication with other types of beings.

4.4 Universals in Vertebrate Vocal Communication

Angela Dassow (Carthage College – Kenosha, US)

License  Creative Commons BY 3.0 Unported license
© Angela Dassow

What is the difference between communication and language? What properties are shared between animal vocalizations and human speech? What approach should be taken when comparing vocalizations produced by different species and how can we employ signal processing techniques to address these questions?

My stimulus talk began by addressing common approaches to studying animal-animal interactions and the attempts made to date to find linguistic properties in animal vocal communication systems. Many such approaches have relied on a single field of interest, either ethology or linguistics. This has resulted in over generalization of capabilities or under appreciation for the complexity of vocal systems. Neither have done a sufficient job to address core evolutionary structures or commonalities across taxa. As an example of this issue, we explored the relevance of searching for vowel harmony in other species. This discussion was based on a study which suggests cotton-top tamarins do not perceive vowel harmony in playback experiments. While there is experimental evidence to support this conclusion, the fundamental question of why would we expect to find vowel harmony in tamarins, when it doesn't exist in all human languages remains. I proposed that this issue and others similar to it, could be avoided by employing an interdisciplinary approach to searching for phonological properties in animal communication systems.

To address the question of what approach should be taken when comparing vocalizations of various species, I proposed a movement from focusing on vocal repertoires, which can be misleading with respect to vocal complexity, to analyzing sequences of acoustic units. As an example of this issue, we discussed the problem of categorizing acoustic units in species with graded vocalizations versus animals with discrete vocalizations. For animals with discrete vocalizations, there is a mismatch between the number of measureable acoustic categories in a given species and their evolutionary relationships. For example, the Irrawaddy dolphin produces fewer categories of sound than the Madagascar treefrog or the Australian long-neck turtle. If we were to use the number of acoustic categories as our main measure of complexity in vocal communication systems, we may erroneously conclude that a dolphin has a more primitive form of communication than a frog or a turtle. To avoid this issue, it is important to examine how these units are being combined. Sequence analysis may offer insight into potential meaning in a vocal communication system which would provide a basis for comparisons made between closely related taxa, such as the sixteen species of extant gibbons.

4.5 Temporal models for machine listening in mixed audio scenes

Dan Stowell (Queen Mary University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Dan Stowell

We wish to be able to analyse soundscapes with multiple vocalising individuals, where those individuals might be human, animal, or otherwise, and might or might not be interacting. Many current models for analysing vocalisation sequences are surprisingly limited for this purpose: they assume there is a single symbol sequence with a strict chain of causality (this might be one individual, or a turn-taking exchange between individuals); they neglect important aspects such as the timing of vocalisations; they assume each vocal unit is a single quantum of meaning.

Simplified models enable efficient inference and can be applied to many species – we do gain a lot from the abstraction of the Markov model, for example. Generic models are essential for handling outdoor sound scenes with dozens of potential species present. But in order to apply machine listening methods to multi-party sound scenes, we need models designed for multiple parties acting in parallel. These models have two complementary

purposes: we fit them to data to measure animal behaviour, and we use fitted models to make inferences in new sound recordings.

I give two specific examples of multi-party models:

1. Multiple Markov renewal processes running in parallel. With this, we can segregate concurrent streams of events. [1]
2. A point-process model in which calls from individuals influence each others' probability of calling. This deals well with multiple parallel influences converging on an individual. With it, we characterise the communication network in a group. [2]

These two paradigms hint at ways forward. Future methods will need richer underlying structure – but what? Sequence modelling? Affective state? Physiological state? Theory of mind? The key question for feasible analysis is, how little can we get away with?

As a separate issue I also discuss “active spaces” and our ideas of signal content. We often treat a vocal unit as having a single purpose and a single audience. In animal communication the concept of an “active space” is the physical space in which a receiver can hear enough of the sound to decode the message conveyed. But it is well known to students of human language that a single utterance can simultaneously have multiple meanings targeted at different audiences. Birdsong contains structural features such as chirp sounds (rapid frequency modulation) which offer a mechanism for multivalent utterances with different spatial extents [3]. We shouldn't accidentally overlook that animals might make use of such possibilities.

References

- 1 D. Stowell and M. D. Plumbley, *Segregating event streams and noise with a Markov renewal process model*. *Journal of Machine Learning Research* 14, 1891–1916, 2013.
- 2 D. Stowell, L. F. Gill, and D. Clayton. *Detailed temporal structure of communication networks in groups of songbirds*. *Journal of the Royal Society Interface*, 13(119), 2016.
- 3 Mathevon, N. and Aubin, T. and Viellard, J. and da Silva, M.-L. and Sebe, F. and Boscolo, D., *Singing in the Rain Forest: How a Tropical Bird Song Transfers Information*. *Plos One* 3(2), 2008.

5 Open Problems

5.1 Statement by Andrey Anikin

Andrey Anikin (Lund University, SE)

License  Creative Commons BY 3.0 Unported license
© Andrey Anikin

The central assumption, for me, is that humans possess a number of species-specific (innate, as opposed to culturally learned) vocalizations, at least some of which are shared with other primates. If we can pinpoint these vocalizations through phylogenetic reconstruction and cross-cultural research, we will have a better understanding of the developmental constraints under which humans acquire their vocal repertoire, including both non-speech sounds and prosodic features of spoken language. This, in turn, will improve human-machine interaction through better recognition and production of vocalizations and prosodically natural speech by machines.

This formulation is broad enough to make it natural to include humans, animals, and robots in the same framework, but still specific enough to lead to testable predictions for

empirical research and to have specific practical implications for affective computing. To unpack, this view of VIHAR involves coordinated efforts and contributions from three fields, as follows:

1. Animal communication.
 - a. Data. To know which sounds humans share with other primates, it is essential to have good descriptions of the vocal repertoire of species closely related to humans, especially the great apes: the acoustic form of vocalizations and typical contexts of their production.
 - b. Method. Researchers studying vocal communication in animals cannot simply ask their subjects what a sound “means”. This has led to a search for stringent methods of classifying sounds into acoustic types (unsupervised classification using some form of cluster analysis, etc), without assuming a priori that each vocalization is specific to one particular context. In my opinion, this methodology is superior compared to the tendency in human research to map sound onto meaning directly, bypassing the level of vocalization.
2. Psychology.
 - a. Large cross-cultural corpora. The existing corpora of non-linguistic vocalizations are relatively small (e.g., compared to the size of speech corpora and collections of animal vocalizations) and limited to a few Western cultures. To find acoustic universals, larger and more diverse corpora have to become available.
 - b. Sound-to-meaning mapping. The relative contribution of within-call and between-call variation needs to be addressed. What range of emotions can a scream indicate? Are the acoustic differences between a scream of anger vs. fear the same as those between an aggressive “evil” laugh and a friendly laugh? Do Morton’s structural-motivational rules apply to human vocalizations? Are (some) vocalizations and/or emotions perceived categorically? These and other questions can be approached via perceptual studies of both natural recorded vocalizations and synthetic sounds (hybrids of natural vocalizations and/or sounds generated “from scratch”).
3. Affective computing.
 - a. Machine learning for sound recognition. This buzzing field is developing very rapidly, but arguably suffers from a piecemeal approach with each team using different training corpora and categories. In my opinion, a more systematic approach with standardized corpora and a more theoretically justified architecture could improve the generalizability of results. In particular, it may be fruitful to introduce a priori constraints on classifiers (e.g., specify dedicated detectors for innate vocalizations, such as laughs and screams) and an intermediate level of acoustic categories distinct from meaning.
 - b. Sound production. There is already considerable interest in producing emotionally charged computer speech. Non-speech vocalizations are a natural extension of this project, and again, just as with recognition, their production can benefit from a more theoretically sound framework. I can conclude by stating, in all humbleness, that I’ve been trying to peck at the problem from all of the perspectives described above. To do more than scratch the surface, however, collaborative efforts are a vital necessity, which is why I believe that VIHAR as a cross-disciplinary framework is the answer.

5.2 Statement by Timo Baumann

Timo Baumann (Universität Hamburg, DE)

License  Creative Commons BY 3.0 Unported license
© Timo Baumann

Highly Responsive Vocal Interaction through Incremental Processing

Vocal interaction (and this is not limited to vocal interaction but also extends to gesture, mimickry and interactive behaviours) is like an intricate dance: what one interlocutor does is potentially immediately analyzed and interpreted by the other and likely incorporated in that interlocutors response behaviour (e.g. backchannelling while listening to speech). While taking turns (a coarse-grained differentiation of sending/receiving in an ongoing interaction) is the *modus operandi* of most human-machine interaction, the more responsive behaviours like backchannelling, blinking, and timing contributions are probably similarly important to achieve good and natural interaction performance.

I work in the area of system architectures for very low-latency reactions and controllable reflexive behaviours, based on incremental processing [1] which allows the concurrent and modular processing of information *as it happens*, including the extrapolation/prediction into the future. Challenges in incremental processing are plentiful and my systems focus on novel interactive behaviours rather than on accomplishing well what existing systems already do (activities like booking a train ticket). Thus, the topic of VIHAR interests me primarily for two reasons:

- Human-animal interaction is often less task-driven and more interaction-driven than human-human interaction (and spoken human-machine interaction which focuses on solving particular problems). Thus, it's a domain in which meaningful interactions are first becoming feasible for incremental systems and I want to learn from researchers on animal interaction about the underlying patterns. Similarly, I believe that contact with roboticists will help to improve interaction capabilities of robots.
- Secondly, I am interested in optimality of the complex interaction system (between and among humans, animals and robots). The wide variety of decision making that is possible at any moment during an interaction and may just look like a small cause may have large effects on the overall outcome. Yet, it is unclear which causes have which effects and to correctly anticipate their magnitudes. I believe VIHAR as a testbed of interaction research is highly valuable to sketch out the possible design spaces of various (natural) interaction systems. What is more, I believe that ultimate human-machine interaction need not necessarily mimic human-human interaction patterns but that better spots in the interaction design space may exist. Inspiration across species-specific research will be very helpful to find better ways of interacting.

References

- 1 Baumann, Timo, *Incremental Spoken Dialogue Processing: Architecture and Lower-level Components*. PhD Thesis, Universität Bielefeld, Germany, 2013.

5.3 Statement by Tony Belpaeme

Tony Belpaeme (University of Plymouth, GB)

License  Creative Commons BY 3.0 Unported license
© Tony Belpaeme

How can robots tap into interactivity?

What fascinates me is the point where vocal interactivity becomes verbal interactivity. The point where vocal utterances are no longer mere grunts, whistles or calls, but where the vocal signal is a package containing distinct chunks. These chunks seem to be the solution (one of many) which animals and humans adopted to communicate symbolic semantics. In animal communication, these chunks seem to break the boundaries of mating, alarm and territorial calls, and carry more complex meaning: they still might be alarm calls, but will now –for example– distinguish the type of threat, as some lemurs and monkeys do. In human language, vocal chunks are now words or grammatical markers, and when strung together they can carry complex, recursive semantic content.

Human cognition relies heavily on intelligent others to evolve and develop, and vocal/verbal communication plays a central role here. When trying to build intelligent machines, such as robots, these not only require the ability to interact with people, but might need a process which helps them tap into human interactions for to mere purpose of bootstrapping and developing their artificial cognition. Concepts, for example, are only to a certain extent acquired through perceiving the physical environment (the so-called “physical grounding” of concepts), but are predominantly subject to a cultural process which relies on interaction. We learn to demarcate the concept of RED not just through experiencing red, but through communicating with others using the word “red” in an appropriate context [1].

Can machines – computers, cloud-based systems, robots – have access to these concepts? To some extent it seems that it is possible to let machines tap into human communication and extract semantic structure: big data approaches show that the simple processes of co-occurrence and correlation can extract semantic relations from mere text. But are there limits to our current methods? Big data and Deep Learning are very much en vogue, and it would seem that the performance of their applications, such as speech recognition, keeps improving with ever more data. But while they are connectionist methods, and therefore have some natural plausibility, they also require huge amounts of annotated data and are therefore fundamentally different to natural learning processes. So a question we need to ask is: are there skills that are fundamentally outside the grasp of these new AI techniques?

One aspect that sets these machine learning methods apart from human learning is the fact that they are batch learners: they feed on huge datasets without the need to interact while learning. Human learning and social learning in animals rely on a tightly coupled interactions between learner and tutor. The tutor, often an adult, will spend considerable resources teaching the learner and will shape the interaction to meet the learner’s needs. From motherese to acquire speech sounds to demonstrations of skills, people seems to have a propensity to structure their interactions to allow the transmission of knowledge and skills. Can machines leverage this to move away from the need for large training sets? Would people be willing to teach machines? What would human-robot interaction look like if machines would learn through interaction? And while building such robots, we not only build novel learning methods, but also develop new methods with which we might study interaction and cognition. We have build a set-up to explore these questions and results indicate that indeed people build a mental model of the robot and tailor the interaction to fit the robot’s learning needs [2].

A different, but still related issue is the relation people have with machines, and with robots in specific. We know that robots are seen as having agency and that much of what a robot does is interpreted as being meaningful. When developing robots to interact with people we need to be aware of how the robot verbal and non-verbal behaviour will be interpreted. With respect to vocal communication, we are quite used to hearing robots utter clicks and beeps, which we call non-linguistic utterances [3]. These utterances are readily interpreted as meaningful by people and seem to be subject to categorical perception, showing how neural mechanisms which evolved for natural communication seems to be sensitive to artificial communicative acts as well [4].

References

- 1 Luc Steels and Tony Belpaeme, *Coordinating perceptually grounded categories through language: a case study for colour*. Behavioral and brain sciences 28(4), 469–488, 2005.
- 2 Joachim de Greeff and Tony Belpaeme, *Why Robots Should Be Social: Enhancing Machine Learning through Social Human-Robot Interaction*. PLOS ONE 10(9): e0138061, 2015
- 3 Selma Yilmazyildiz, Robin Read, Tony Belpaeme, and Werner Verhelst, *Review of Semantic-Free Utterances in Social Human-Robot Interaction*. International Journal Of Human-Computer Interaction, 32(1), 2016
- 4 Robin Read and Tony Belpaeme, *People Interpret Robotic Non-linguistic Utterances Categorically*. International Journal of Social Robotics, 8(1), 31–50, 2016

5.4 Statement by Elodie Briefer

Elodie Briefer (ETH Zürich, CH)

License  Creative Commons BY 3.0 Unported license
© Elodie Briefer

Since my PhD, I have been investigating the acoustic communication of several species, including skylarks, fallow deer, goats, horses, Przewalski's horses, pigs, wild boars and cattle. All these species differ widely in the form and complexity of the sounds they produce and raise different questions/challenges for VIHAR. I will here only focus only on my main current project, the study of vocal expression and contagion of emotions in ungulates.

Emotions play an important role in social species, because they guide behavioural decisions in response to events or stimuli of importance for the organism and hence regulate social interactions (e.g. approach or avoidance). Indicators of emotions in human voice have been studied in detail. However, similar studies testing a direct link between emotions and vocal structure in non-human animals are rare. In particular, little is known about how animals encode in their vocalisations, information about the valence (positive/negative) of the emotion they are experiencing. Furthermore, the potential for emotions to be transmitted to conspecifics and hetero-specifics through vocalisations (vocal contagion of emotions) has been poorly studied. A comparative approach between humans and other animals would give us a better understanding of how the expression of emotions evolved.

My current project aims at combining methods to study emotions and vocalisations in order to investigate the evolution of vocal expression of emotions and the impact of domestication on human-animal communication of emotions. My project focusses on (1) vocal expression of emotions in domestic and wild ungulates; (2) perception and contagion of emotions between conspecifics; (3) perception and contagion of emotions between closely related domestic and wild ungulates; (4) perception and contagion of emotions between

domestic and wild ungulates and humans. It includes goats, horses, Przewalski's horses, pigs, wild boars and cattle.

I am listing below some of the challenges for VIHAR that my project raises:

Fundamental challenges

- How can we best compare vocal expression of emotions in animals and humans? My research focusses on “subtle” acoustic variation occurring within call types (e.g. within horse whinnies) as a function of emotional valence and arousal; Is it correct to compare emotion-related changes in vocalisation types (e.g. bark → growl) to human nonverbal emotion expressions (e.g. laughter → screams), while variation within vocalisation types is closer to affective prosody?
- Can we really differentiate between “emotional” and “intentional” signals in animals?

The main application of my research resides in the assessment and improvement of animal welfare.

Applications

- Emotion expression: Development of automated tools that would recognize animal's emotions from their vocalisations. Can these tools be trained on the calls of each individual? Such tools could allow animal keepers to be informed when a certain threshold of vocalisations indicating negative emotions are produced and could thus take action to improve welfare.
- Emotional contagion: Development of acoustic tools that would decrease negative arousal (e.g. during stressful husbandry procedure) and promote positive emotions. These tools could take the form of synthetic vocalisations based on our knowledge of parameters that trigger emotions in receivers.

5.5 Statement by Nick Campbell

Nick Campbell (Trinity College Dublin, IE)

License  Creative Commons BY 3.0 Unported license
© Nick Campbell

Pragmatism, Context-sensitivity, and the Robot-Dialogue Interface

I come to this meeting from a background of speech processing for human-human translation machines with a specific emphasis on speech synthesis, particularly concerning utterance generation and timing. Now working on autonomous robot dialogue interfaces for human-robot interaction, my prime interest is in the style and content of utterances delivered by the device: for a natural-seeming spoken interaction, the speech must be relaxed, apparently spontaneous, and contextually appropriate.

Previous research with the “Herme” conversational robot [1] has shown that even without an understanding module (or even functioning speech recognition) a machine (robot) can maintain a natural-seeming conversation with a human for between three and five minutes. However, going beyond this simple time limit will require an element of understanding on the part of the robot in order to continue the conversation and contribute satisfactorily.

The goal of our work in the Speech Communication Lab at the University of Dublin, Trinity College, is not to create yet another chatbot but to understand how to improve the delivery of predetermined utterances in the context of engaging the interlocutor and assessing the

cognitive effect of each message. For me, an issue to be addressed at this VIHAR meeting is the extent of understanding required by the robot for an efficient situated dialogue; whether full Theory-of-Mind is required for linguistic grounding or whether simpler pragmatic/functional constraints of the dialogue context can sufficiently restrict the interaction for the robot to respond from a limited list of pre-prepared default utterances or utterance-types. I bring to the meeting a small study of dog barks in the context of human engagement and show from that work how rather than relying on an underlying ‘language of barks’ that each dog/human pair has to learn, there is a situational context dependency from which an interpretation can be gained.

In the barking study [2], we did not find that bark type generalised widely between different dogs of the same species but infer that each dog had developed its own similar-sounding bark type in response to a common set of everyday situations under common articulatory constraints. In the context of human-robot interaction, it may prove to be the case that rather than share a common (human) language, sufficient sounds may trigger an appropriate reaction in a given context when the constraints of that context are understood by both participants. In implementing such a model, we focus first on determining the degree of engagement of the human (i.e., where his or her attention is directed when the robot is speaking or about to speak) and on maintaining sufficient contact throughout a speech interlude so that the desired message may be delivered and the interaction satisfactorily completed.

Here the example of a receptionist robot dialogue interface collaboratively built during the recent eINTERFACE workshop becomes relevant [3]; the situation is extremely constrained but practical, and the receptionist simply has to direct each customer/patient to the desired room as they arrive. In our test case, there are only two rooms and only two humans in the robot’s universe. We built an exhaustive model of how to deal with each customer (including the utterances required for each move) and how to manage the queueing of customers when more than one was present. The robot also had a set of idling behaviours to return to when each customer was served. Rather than program this behaviour deterministically, we had access to the Flipper dialogue management engine [4] that continually tests the environment for a set of given conditions and then acts (and resets the environmental state) accordingly. The set of condition-behaviour-response tokens is large but finite. The success of this model depends on the responses also being finite, but we claim that this might be the case for a large number of real-world situations and that full ‘understanding’, particularly ‘linguistic understanding’ on the part of the robot, might not be necessary. The use of other sensors, however, is mandatory, and our robot is able to see the environment and to recognise simple gestures such as pointing.

It will be interesting to hear whether colleagues from the animal sciences have any contributions to make to this model from their observations of animal behaviour and of the very restricted use of ‘language’ that animals appear to make.

References

- 1 Han, J., Gilmartin, E., De Looze, C., Vaughan, B., and Campbell, N., *The Herme Database of Spontaneous Multimodal Human–Robot Dialogues.*, Conference on Language Resources and Evaluation (LREC’12), Istanbul, Turkey, 21-27 May 2012, edited by ELRA, 2012
- 2 Nick Campbell, *An acoustic analysis of 7395 dog barks.* in editor(s) Rudiger Hoffman, Festschrift, Book Chapter, 2013
- 3 Daniel Davison, Binnur Gorer, Jan Kolkmeier, Jeroen Linssen, Bob Schadenberg, Bob van de Vijver, Nick Campbell, Edwin Dertien, Dennis Reidsma *Things that Make Robots*

- Go HMMM: Heterogeneous Multilevel Multimodal Mixing to Realise Fluent, Multiparty, Human-Robot Interaction.* Proc eNTERFACE, Enschede, Netherlands, 2016 (in press)
- 4 MarkMaat, T., and Heylen, D., *Flipper: An Information State Component for Spoken Dialogue Systems.* in Intelligent Virtual Agents. Reykjavik: Springer Verlag, 2011, pp. 470–472.

5.6 Statement by Fred Cummins

Fred Cummins (University College Dublin, IE)

License  Creative Commons BY 3.0 Unported license
© Fred Cummins

My work thematises *Joint Speech*, defined as speech produced by multiple people at the same time. It is a familiar form of speech, serving to empirically pick out highly valued domains of human practice including practices of prayer, ritual, protest, and the enactment of collective identity among sports fans. The absence of any empirical scientific work in this domain is revealing. It demonstrates a fixation within the human sciences on a solipsistic, Cartesian, and, yes, very obviously Christian, approach to the person, that has treated language as a form of modality-neutral passing of encoded messages containing information. This obsession of the science of “language” is necessarily blind to the manner in which we bring a shared world into being through our coordinated practices, including our vocalisations. In my view, linguistics has failed to identify language in the first place.

The study of joint speech serves to bring some neglected themes to the fore in place of the concerns of academic linguistics. Joint speech is clearly a highly central example of language use, as old as humanity, and instrumental in bringing many kinds of human collectivities into being. Yet in joint speech some familiar distinctions vanish. The opposition of speaker and listener is no longer relevant, as everybody is both and the texts uttered are authored elsewhere. Likewise, there is no principled manner to distinguish between speech and music any more, as we find all possible points on a continuum from the amusical recitation of an oath on a singular occasion, through the rhythmically and melodically exaggerated chants of repeated prayers or protest calls, to the unison singing of plainsong or the familiar Happy Birthday. Joint speech cannot be replaced by written texts. It is performative at its core: taking part in joint speech practices is not a neutral activity conducted in an intellectual tone. It is an act of commitment, through which many of the structural elements of any human society are made manifest and are maintained by doing. Studying joint speech changes the principal themes we might pursue as we study vocal behaviour, and these concerns, once recognised, may be extended far beyond the rather narrow specification of the definition of joint speech itself. They appear to me to shed light on all human vocal communication and to extend naturally to human-animal interactions as well. To the extent that they generate novel ways of conceiving of the role of the voice in interaction, they may prove relevant to human-robot interactions as well. I have not pursued that particular thread yet.

Once the message-passing metaphor is no longer relevant, we uncover instead a realisation that the *real time recurrent interaction* among participants is an essential element to any joint speech event. Participants are in contact with each other in a very important manner. This recognition also serves to shift the focus from notions of representation and reference, to an awareness of the importance of *co-presence* among participants. Making *vocal interactivity* a research theme seems to me to offer a more promising starting point for understanding

what people are doing in such situations than anything on offer within an anaemic and abstract “linguistics”.

Occasions in which joint speech is important are inevitably embedded in rich and highly charged suties of practices that are themselves highly informative about the values and lifeworlds of the participants. Any single instance of joint speech needs to be interpreted with a keen sense of the context in which it is embedded. Transcription is irrelevant. What is needed instead is the notion of *thick description*, providing as much supporting material to reveal the specific context-bound manner in which one or other instance of joint speech is integrated into meaning-making activities. I suspect we have a lot more observation to do before we can redress the shortcomings of the received syntax-first approach to language that seems to be good mainly for bible translation.

References

- 1 Cummins, F. (2014) *Voice, (inter-)subjectivity, and real-time recurrent interaction* *Frontiers in Psychology: Cognitive Science*, 5(760).
- 2 Cummins, F. (2014) *The remarkable unremarkableness of joint speech* in *Proceedings of the 10th International Seminar on Speech Production*, pages 73–77, Cologne, DE.

5.7 Statement by Angela Dassow

Angela Dassow (Carthage College – Kenosha, US)

License  Creative Commons BY 3.0 Unported license
© Angela Dassow

Understanding evolutionary relationships through examining vocal interactivity

In non-human animals, communication is widely viewed as a behavior; it is a reflexive activity designed to produce behavioral responses in conspecifics or across species. In contrast, language is a human affair. It transfers conceptual knowledge from speaker to listener and has extraordinarily generalizable descriptive powers. While spoken language may lead to behavior, this is unnecessary.

Longstanding debates regarding the use of language in non-human animals have focused on nested and recursive syntactic structures as proxies for the core competencies associated with human cognition. However, current understanding of non-human cognition precludes the need for complex conceptual representations that require the deep mathematical structure of human language. Put plainly, what precisely do these animals have to think about, let alone communicate? Current paucity in understanding cognition in other species renders this question simply provocative.

That aside, evaluating the existence of language in non-humans solely via analogies to syntax begs the question: how would vocalizations hypothetically possessing syntax be constructed? While insistence of parity with the structural complexity of human language are widespread, the precursor questions of morphology and phonology in non-humans have largely been ignored. The constituent pieces that enable formation of more elaborate structures must be examined before comparisons of structural complexity can be met.

The goal of my research is to characterize linguistic commonalities in different vertebrate species that communicate vocally. Specifically, my interests lie within the following problems:

- **Commonalities derived thru evolution:** Much like Darwin noticed similarities of physical features such as wings, we are noticing similarities in sound categories across several diverse species. As a component of evolution, selection can only occur on existing

structures. This stands to reason then for vocal species, that there is some connection to how vocalizations are made as well as, why and what meaning the vocalizations have. Part of my research agenda is to better understand these connections.

- **Developmental differences within clades:** Within monophyletic groups, there is variation between different species vocal development patterns. While vocal production and comprehension is innate in some species, other species require a sensorimotor learning style and others require something in between. I am developing methods to make inferences of vocal complexity based upon a pre-existing understanding of how various species learn to meaningfully vocalize. My goal in this pursuit is to determine what environmental and genetic factors are important for developing a more complex way of communicating vocally.
- **Potential for linguistic structure:** Cognitive studies of animals have provided some insight into what certain species are capable of. My work strives to further this understanding by viewing this problem from a cross-disciplinary approach. Instead of focusing on making a direct connection to human language, I first examine what meaningful connections individuals within single species are making with conspecifics. Once these connections are explored, I then expand my view of the interactions to include individuals from other species that may come into regular contact with my focal species. My goal in this approach is to first understand what meaningful communication may be going on within a community with which the species has coevolved in before making a larger leap towards how that may relate to us.

5.8 Statement by Robert Eklund

Robert Eklund (Linköping University, SE)

License  Creative Commons BY 3.0 Unported license
© Robert Eklund

Personal statement

Given a background in Speech Technology (I worked on the first concatenative speech synthesizer for Swedish, the first commercial ASR system for Swedish (now Nuance) and the first open prompt human–computer support system in Scandinavia (Telia 90 200) it has, for a long time been “natural” for me to think in terms of interaction, and concepts like agents, avatars, Theory of Mind and interface design (auditory and visual) have all been part of parcel of my work activities during the period 1994 to (roughly) 2012.

For completely unrelated reasons I started to expand my research interests into animal vocalizations in the year 2009 when I made a recording of a cheetah purring ¹ and these activities did then snowball into a five-year-long project where me and colleagues will study human–cat interaction with focus on prosody/melodic aspects ².

My Stimulus Talk during the Dagstuhl conference did not focus on or describe my previous research on the topic (cheetah, lion and domestic cat vocalizations) but instead raised some “larger issues” concerning “cross-species” (with a wide definition of ‘species’, including robots) communication.

These will shortly be described below:

¹ <http://www.youtube.com/watch?v=ZFvULxbN3NM>

² <http://meowsic.info>

Personality issues

The literature is replete with studies of personality (and was crucial in e.g. how to put together submarine crews during WWII). However, such studies are not constrained to human but several studies of personality in different species of felids are also to be found (see Bibliography), partly for husbandry reasons. My issue-to-raise here is to what extent individual personalities play a role when humans interact with other species.

New form of “uncanny valley”?

In 1970 Masahiro Mori published a paper title “The Uncanny Valley” (in Japanese translation) [1] where he described a dip in the easiness with which we approach and regard humanoids. If these are completely not like us (like 1930s teddy bears or cartoon characters) we have no problem, which is also the case if there are very similar to us. However, if something is “eerily” similar to us – not completely not like us, but not completely not like us, either – we get a spooky feeling around them. My question here is whether this can occur in the auditory domain, too. If computers sound very much like machines, or whether animals respond to or signal to us, in ways that are definitely not human-like, we (obviously) have no problem. But what happens when either robots or animals start communicate with us in very human-like manners – both voice-quality and content-wise: will this created another/a new form of more abstract uncanny valley?

Was Wittgenstein right?

Wittgenstein famously stated that “if a lion could speak we would not understand him”. This obviously played on the idea that the lion world is so basically different from the human world that there is no way that we could understand the lion’s worldview. (Note that this argument has also been forwarded within anthropology when studying other – most often non-Western cultures.) But is this necessarily true? Although undeniably true that a lot has happened since humans lived “on the savannah”, we still most likely share the same basic emotions, and are governed by them. This should, in my view, provide some solid common ground for mutual understanding.

Health effects?

To spend time with a pet, or even robots, is beneficial from a health perspective. Will this effect be enhanced by improved communication with animals or robots? Or will a potential new uncanny valley effect reverse this?

Symbol mapping?

The cheetah is particularly famous for its agonistic moan-growl-hiss-spit+paw hit sequence³ (most felids exhibit this, minus the paw hit). How to interpret this sequence? As one agonistic sequence that qualitatively changes character as it escalates, or as four different “symbols”, all with their own intrinsic meaning? The basic question is: to what extent can we use the standard linguistic toolbox when we describe animal vocalizations?

³ <http://www.youtube.com/watch?v=bBIf5g2Fp1U&feature=youtu.be>

Language learning?

That several species of animals are capable of language learning – and consequently also dialectal variation – has been known since Aristotle [3]. What can we learn about our own acquisition of language, phylogenetically, from the study of language learning in animals?

Role of hearing?

Animals vary a lot when it comes to hearing abilities, both frequency-wise and from a source location point of view (see Bibliography below). To what extent do we need to take other species' hearing abilities into account when trying to communicate across species? Case in point: the Beluga whale described in [2] who deliberately made an effort to vocalize outside its comfort zone when addressing humans.

Motherese?

It is well-known that humans – at least in the western world – make use of what is sometimes called “motherese” when they address infants (or small children). This speech style is characterized by an exaggerated prosody and simplified phone and word repertoires. It is also known that humans use the same “trick” when addressing their pet animals. Does this have any benefits on the animal side of things, or is it simply something that we do semi-automatically for our own benefit?

Summing it all up

There are, obviously, loads of things to consider when expanding our knowledge on how animals communicate, and on how we as humans can improve our communication with those animals. Although not exactly the same, there is considerable overlap in our communication with robots (and animated agents and/or avatars) and there is no doubt in my mind that there will be vast cross-fertilization between all those fields in the future. And I hope to be part of this!

Web resources

- <http://roberteklund.info>
- <http://ingressivespeech.info>
- <http://purring.info>
- <http://meowsic.info>

References

- 1 Mori, Masahiro, *The Uncanny Valley*. Energy vol 7, no 4, 33–35, 1970.
- 2 Ridgway, Sam, Donald Carders, Michelle Jeffries & Mark Todds, *Spontaneous human speech mimicry by a cetacean*. Current Biology, vol 22, no 20, R860–R861, 2012.
- 3 Zirin, Ronald A., *Aristotle's Biology of Language*. Transactions of the American Philological Association, vol 110, 325–347, 1980.

5.9 Statement by Julie E. Elie

Julie E. Elie (University of California – Berkeley, US)

License  Creative Commons BY 3.0 Unported license
© Julie E. Elie

As humans, spoken language is central in our everyday life. We use it to exchange information, to express our emotions and to form social bonds with other human beings. The auditory system plays a fundamental role in the perception and interpretation of these communication sounds. Both in humans and animals, the auditory system parses the auditory stream coming to the ear and extracts the behaviorally relevant acoustic features of sounds, leading to the percept of meaning for communication signals. Auditory neuroscientists have obtained a relatively good model of how complex sounds are represented in the primary auditory cortex primarily in terms of their spectro-temporal features. We also know that a network of higher-level auditory and associative cortical areas is involved in processing speech in humans and communication calls in animals. However, the neural circuits and the corresponding non-linear transformations that occur between primary auditory cortical areas and cortical regions that categorize communication sounds in terms of their meaning remains unknown. One first area of knowledge that I think needs a research effort is to identify the computational steps leading from the perception of communicative sounds to the invariant representation of meaning in the brain.

Furthermore, as young humans, we don't only learn to produce speech but also learn to understand the meaning of words and other non-verbal vocal communication signals. The correct interpretation of communication signals is necessary not only for eliciting the appropriate behavioral response but also for learning the appropriate usage of the vocalization. This ability to learn the meaning of vocalizations is not restricted to humans. Young vervet monkeys, for instance, progressively refine their reaction to alarm calls, adopting progressively the right behavioral response to the nature of the predator (e.g. terrestrial or aerial) signaled in the alarm call. While the neural basis of vocal learning and plasticity has been well studied in animal models, mostly in songbirds, the role of learning and its neural underpinnings in the correct interpretation of communication signals has yet to be investigated. Previous research has demonstrated that exposition to particular sound statistics or reinforcement learning with sounds does enhance the neural representations of these behaviorally relevant sounds. However, the role of plasticity in auditory cortex during development for the correct categorization of communication signal is unknown. While the auditory extraction of some relevant behavioral information could be innately implemented in the wiring of the brain (such as the basic response to alarm vocalizations in vervet monkeys), the extraction of other informative features (such as the type of predator encoded in the alarm call) is likely learned by experience and likely rely on the maturation of the auditory cortex. As such, another area that is likely of interest is to explore the extent of innate processing for social cues in vocalizations and to describe changes in neural processing of social information as the brain matures. Knowing the maturation/learning steps in animals might help us better calibrate machines/robots that should also be able to mature/learn with the environment they are navigating in.

Finally, besides the meaning conveyed by single sound elements or sequences of signals, the rhythm with which individuals exchange information might also be informative about the vocalizer internal/emotional state, or about the urgency of the situation. Brief alarm calls repeated several times might for instance be more effective in achieving the desired behavior of rising others attention and fleeing for cover and could gradually indicate the

imminence of the danger. In another line of studies, the synchrony of vocal exchanges between close related individuals seem also to serve social relationship by maintaining or straightening the bonds. Duets between paired individuals in the context of territory defense is as such most likely advertising the strength of the alliance between the mates to potential intruders. When such duets are performed in a more intimate context then the hypothesis of the pair-bond reinforcement has been proposed. However, we still don't know in terms of both physiology and information content, what are the consequences of the precise synchrony between individuals during these vocal interactions, and we are even further from understanding how this precise timing is achieved.

5.10 Statement by Sabrina Engesser

Sabrina Engesser (Universität Zürich, CH)

License © Creative Commons BY 3.0 Unported license
© Sabrina Engesser

Vocal combinations in non-human animals

Research over the last five decades has indicated that numerous aspects of human language also exist in non-human communication systems [1]. Reference and intentionality represent two key components of language, with meaning being assigned to vocal structures, and information being voluntarily communicated [2]. Analogue forms of these components are found in various forms in non-human species. Animal vocalisations can, for example, refer to current external events or objects [3], and signals can be flexibly used by animals to inform or manipulate receivers, or equally, information can be withheld in the presence or absence of certain individuals [4]. Such strategic, flexible use of vocalisations indicates that vocalisations and the decision to call are not necessarily hardwired in animals, but individuals might have a certain degree of control over their vocal production [5]. Whilst these findings have been argued to provide insights into understanding the evolution of linguistic abilities central to language, there remains a problem with regard to language's generative nature, particularly its evolutionary origin and the selective conditions promoting its emergence [1, 6]. Theoretical work hypothesises that language's combinatorial layers evolved in order to overcome productional and perceptual limitations [7]. Specifically, stringing meaningless sounds (phonemes) together can enhance the discriminability between otherwise similar sounding signals, and hence decrease perception mistakes [7]. Once the number of messages to be encoded exceeds the number of discrete signals present in a communicative system, and in order to offset memory limitations, meaningful signals can then be assembled in a systematic way into higher order meaningful structures [7].

Empirical data on animal communication systems can help to test such hypotheses, and a broad comparative approach can provide insights into the evolutionary progression of human language's combinatorial components. In line with the comparative approach, my research investigates the prevalence and diversity of vocal combinations in two highly social passerine birds which do not sing, but instead possess an array of discrete vocalisations: southern pied babblers (*Turdoides bicolor*) and chestnut-crowned babblers (*Pomatostomus ruficeps*). Given the extensive array of behaviours that require coordination, there has likely been a significant selective pressure on both species to evolve new and diverse call types. However, like most animal species, babblers appear to be anatomically constrained in the number of different calls they can produce. Combining existing sounds and calls may therefore represent

a potential mechanism applied by both species to increase the amount of information that can be encoded, facilitating the smooth management of a plethora of behaviours upon which the stability of these species' social and breeding system depend (for more information see [8, 9]).

VIHAR related statements/thoughts

While “vocal learning is thought to be a key precursor of [...] language” [10] a fundamental question arises concerning why – of the few known animal taxon possessing the ability to generate novel sounds – this capacity is primarily allocated to the creation of sound combinations devoid of conventional meaning [11], with complex structures/songs being primarily driven by female preferences for elaborate male songs, or by selection for individually recognisable signals functioning, for example, in bonding behaviour [12, 13]. Potentially the loose association between signal structure and conventional meaning has enabled the creation of ever-more complex vocal sequences. But how crucial is vocal learning for the evolution of meaningful generative capacities (i.e. rudimentary phonemic and syntactic structures)?

Whilst “the physical apparatus for articulation and audition differs from species to species” [10], it is crucial to also consider to what degree the environment a species inhabits shapes the spectral features of its vocalisations and their perception. Are species with 'fixed' vocal repertoires actually constrained in their vocal production (i.e. did a species adapt) or do they simply 'adjust' to an 'acoustic/environmental niche' with underlying vocal plasticity? How and to what degree do anatomical and environmental constraints affect the structure of vocal signals, and how does this in turn shape the emergence and the forms of combinatorial structures in non-human animals?

“Vocal interactivity is likely often teleological and is thus conditioned on underlying intentions. [...] To what extent are vocal signals teleological, and is it possible to distinguish between intentional and unintentional vocalisations?” [10]. Besides asking whether a signal is intentional or unintentional, from a receiver's perspective a signal may not necessarily have to be teleological to serve a communicative purpose. Some calls may simply encode emotional states of the caller and still transfer information triggering an evolutionary adaptive response in receivers. Concerning vocal sequences, is intentionality a prerequisite for combinatoriality? Do signals have to be purposefully combined to encode information in a compositional fashion and to be meaningful for receivers?

References

- 1 Hauser M. D., Chomsky N., Fitch W. T. *The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?* Science 298, 2002.
- 2 Hockett C. F.. *The Origin of Speech* Sci. Am. 203, 1960.
- 3 Townsend S. W., Manser M. B.. *Functionally referential communication in mammals: the past, present and the future* Ethology 119, 2013.
- 4 Tomasello M. *Origins of Human Communication* Cambridge, MA: MIT Press, 2008.
- 5 Marler P., Dufty A., Pickert R.. *Vocal communication in the domestic chicken: II. Is a sender sensitive to the presence and nature of a receiver?* Anim. Behav. 34, 1986.
- 6 Bolhuis J. J., Tattersall I., Chomsky N., Berwick R. C. *How Could Language Have Evolved?* PLoS Biol. 12, 2014.
- 7 Nowak M. A., Krakauer D. C. *The evolution of language* Proc. Natl. Acad. Sci. USA 96, 1999.
- 8 Engesser S., Crane J. M., Savage J. L., Russell A. F., Townsend S. W.. *Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System* PLoS Biol. 13, 2015.

- 9 Engesser S., Ridley A. R., Townsend SW. *Meaningful call combinations and compositional processing in the southern pied babbler* Proc. Natl. Acad. Sci. USA 113, 2016.
- 10 Moore R.K., Marxer R., Thill S. *Vocal Interactivity in-and-between Humans, Animals, and Robots* Front. Robot. AI 3, 2016.
- 11 Rendall D. *Q&A: Cognitive ethology – inside the minds of other species* BMC Biol. 11, 2013.
- 12 Catchpole C.K. *Bird song, sexual selection and female choice* Trends Ecol. Evol. 2, 1987.
- 13 Janik V., Slater P. *Vocal Learning in Mammals* Adv. Stud. Behav. 26, 1997.

5.11 Statement by Sarah Hawkins

Sarah Hawkins (University of Cambridge, GB)

License © Creative Commons BY 3.0 Unported license
© Sarah Hawkins

Sound and meaning. Researchers into human’s speech perception typically rely heavily, and often entirely, on the units of formal linguistic theory as the elements that ‘need identifying’, unless the focus is emotion. I believe that privileging such atomistic, non-redundant units that neglect communicative function provides a distorted view of how people understand spoken language: non-redundancy is biologically implausible; information in spectrotemporal properties of the spoken signal is ignored unless it contributes directly to lexical identification and narrow sentence meaning, which leads to unlikely models of perceptual processes; and important aspects of human communication are neglected. An utterance’s meaning can be quite different from the meaning of its individual words and grammar, being modulated by voice quality, facial expression, and the situation itself – cf. the range of responses that hold on or take a break/brake or even the cat’s over there! invite, given different renditions and situations. Rich, subtle meaning can also be conveyed without words yet phonetically reflect the implied words (Hawkins, 2003, Table 2 erratum). So prioritizing linguistic unit identification provides an incomplete ‘sterile world’ analysis – it neither uses all information inherent in the multi-modal physical signal, nor guarantees a full description of the talker’s meaning, nor allows for that intended meaning to be filtered through the listener’s preconceptions. Instead, we need to prioritize the input, rich interpretation, and their interaction. This entails using stimuli that are recorded and responded to in contexts that demand attention to broad meaning, as well as refocussing effort onto the input signal (‘below’ linguistic units), and on how to represent meanings without being forced to depend on identification of intermediary linguistic units. This alternative way to conceptualize speech perception processes may connect more straightforwardly with both animal work and robotics, for when communicative function is clear, meanings can be clearly conveyed without phonetically segmentable units in the physical signal, and I speculate that the processes that make this possible are those that are fundamental to communication within and between species, and also to engaging with inanimate events.

I work with Polysp (POLYsystemic Speech Perception), which centres on mapping perceived properties of the physical signal to metrical and rhythmic structures appropriate for situated communication in the specific language. Details of structures must be species- and language-specific, but in general, sound chunks and associated information (visual, situational) activate competing structures to differing degrees. An attribute of a sound chunk can signify several types of structural element, and different attributes can activate one element. Strongly-activated metrical structures influence mapping by changing weights on

specific sound chunks, depending on the likelihood of one meaning over another and prior knowledge of expected sound patterns in the context. E.g. English /s z m n/ vary acoustically less than /t d/ and ‘th’ as in this, but all are affected by grammatical status. So prediction influences how physical features are attended to, interpreted and hence mapped. When meaning is reached without identifying less-certain elements, these are ‘filled in’ afterwards by pattern completion processes. But if a filled-in candidate structure does not match the perceived rhythm, that structure is discarded. Distinctive (re psycholinguistics) aspects of Polysp include that relative timing is fundamental, no unit can be described independently of its context, identifying units between sound and meaning is not essential, and the distinction between ‘top-down’ and ‘bottom-up’ processes has limited value.

Though its principles have been used for text-to-speech, Polysp has not been implemented as a recognition system, and drawbacks include that it lacks well-specified high-level functional/intentional information capable of dealing efficiently with the detail, and it needs extending to account for interaction. Several perception-action robotics systems have such high-level control, but their speech models do not exploit the rich communicative information available from phonetic detail. I hope the two approaches can inform each other and hence come together.

Interaction & Generality. The claim that rhythm is fundamental to understanding an utterance leads naturally to examining interaction, with the literature suggesting temporal entrainment between musicians, and phase-locked neural oscillations between talkers. We have recent evidence of entrainment across turn boundaries in well-formed Question-Answer pairs, of seamless transfer of pulse between conversational speech and jointly-improvised music, especially when the musical rhythmic pulse is less variable, and new data tentatively suggesting that experience in predicting turn-taking during improvisational music-making and language games enhances empathy amongst teenagers, compared with just playing and (e.g.) rapping together. The questions I ask on slide 3 about interaction and generality seem to me ideally answered in a cross-disciplinary and cross-species forum. I value most highly models that are not just biologically plausible but are also as biologically general as possible. I welcome work that is cautious about making speech and language too special: true, many language attributes seem to be largely specific to humans, but each species’ communication has unique aspects, and for me there is much interest in finding commonalities.

References

- 1 Hawkins, S. (2003) *Roles and representations of systematic fine phonetic detail in speech understanding*. *Journal of Phonetics* 31(3–4), 373–405. <http://dx.doi.org/10.1016/j.wocn.2003.09.006>. Erratum in *J. Phonetics* 32(2), 289.
- 2 Ogden, R., & Hawkins, S. (2015) *Entrainment as a basis for co-ordinated actions in speech*. The Scottish Consortium for ICPHS 2015 (Ed.), 18th International Congress of Phonetic Sciences. Univ. Glasgow; ISBN 978-0-85261-941-4. Paper number 0599.
- 3 Hawkins, S. (2014) *Situational influences on rhythmicity in speech, music, and their interaction*. In R. Smith, T. Rathcke, F. Cummins, K. Overy, S. Scott (eds.) *Communicative Rhythms in Brain and Behaviour*. London: Philosophical Transactions of the Royal Society B 369: 20130398. <http://dx.doi.org/10.1098/rstb.2013.0398>

5.12 Statement by Ricard Marxer

Ricard Marxer (University of Sheffield, GB)

License  Creative Commons BY 3.0 Unported license
© Ricard Marxer

Vocal interaction plays a fundamental role in our day-to-day relations to our environment and to others. We are capable of explaining complex ideas to others and recognising the emotional state of someone from the tone of their voice. Animals make extensive usage of vocalisations, whether to establish territory, sound an alarm or establish social bonding. Vocal signals are also central in the study and design of autonomous agents, nowadays we can perform Internet searches with spoken commands and maintain short conversations with virtual personal assistants.

The research I have conducted has mainly revolved around humans' perception and production of acoustic signals. From the study of music perception and singing voice to the modelling of speech intelligibility, the work I have done investigated several aspects of human listening [1, 2] and multimodal vocal interaction [3].

My interest in organising the VIHAR seminar comes from wanting to understand the underlying principles, commonalities and differences governing vocal interactivity in humans in animals. I'm also interested in seeing how these principles can be used to modify our interactive experiences with autonomous agents, and how these agents could be used to further explore animal behaviour.

I also think understanding the underlying principles of vocal interactivity across species could have an important impact on well-established fields. Knowledge about the formation of perceptual acoustic units throughout species could have an impact on speech recognition systems for under-resourced languages. Better understanding of the role of top-down and bottom-up processes in the perception of acoustic categories could significantly improve human speech intelligibility modelling. Insights on the role of expectations in vocal interaction in both animals and humans could change the way in which we build dialog models or interactive music systems.

In particular some of the initial questions that spark my interest are:

- Are there common processes involved in the categorical perception of acoustic units in humans and animals? What computational models could be used to reproduce such processes?
- What top-down processes are involved in the perceptions of vocal signals in animals? How are these related to those involved in human speech perception?
- What's the role of expectations in the perception of vocal signals (and sequences of them) in both humans and animals? And how may these be exploited when establishing interaction with autonomous agents?
- What principles underlie the perception/production of sequences of acoustic units in animals? How do these relate to the principles governing human vocal interaction?
- What modelling and machine learning techniques can help us understand the general principles (if any) of vocal interactivity across multiple species?
- How do multiple modalities influence the process of acoustic perception in animals? Is there an analogue to the McGurk effect in animal vocalisations?

References

- 1 Marxer R. & Purwins H., *Unsupervised Incremental Online Learning and Prediction of Musical Audio Signals*, in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 5, pp. 863–874, May 2016.

- 2 Marxer R., Barker J., Cooke M. & Garcia Lecumberri M. L., *A corpus of noise-induced word misperceptions for English*, in *J. Acoust. Soc. Am.* (in press), 2016.
- 3 Abel A., Marxer R., Barker J., Watt R., Whitmer B., Derleth P., & Hussain A. A. *A Data Driven Approach to Audiovisual Speech Mapping*, in *Advances in Proc. of BICS 2016*, Beijing, China, November 28-30, 2016.

5.13 Statement by Roger K. Moore

Roger K. Moore (University of Sheffield, GB)

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

Since joining the Speech and Hearing Research (SPandH) group at Sheffield in 2004, I've developed (and published) a unified theory of spoken language processing called PRESENCE (PREdictive SENsorimotor Control and Emulation) which weaves together accounts from a wide variety of different disciplines concerned with the behaviour of living systems in general – many of them outside the normal realms of spoken language – and compiles them into a novel framework that is breathing life into a new generation of research into spoken language processing.

PRESENCE (first published in 2007) presents a number of practical implications with regard to new models for automatic speech recognition and generation. However, it also poses some fundamental questions about the nature of vocal interactivity – not just about speech communication between one human being and another, or between human beings and machines, but questions concerning the foundations on which all such interactive behaviours are based – questions such as:

- How do (living) systems coordinate their activities by vocal signalling?
- What is the role of prosody and emotion in establishing relations between equal/unequal social partners?
- How does mimicry and imitation facilitate learning (e.g. in development)?
- How are interaction skills acquired?
- What evolutionary constraints are implicit in such behaviours?

Contemporary approaches to spoken language interaction and dialogue (e.g. Siri – Apple's voice-enabled personal assistant) are understandably based on rather naïve models of message passing and strict turn-taking. PRESENCE, on the other hand, points to a more fluid model of interactivity based on continuous interaction between coupled dynamical systems. PRESENCE also shows how behaviours such as emotion serve to drive an organism's intentions and that 'empathy' between interacting agents facilitates signalling efficiencies. What I'm trying to do now is to go back and address some of these fundamental scientific and technical questions, and what seems to be needed is a comparative approach that is not limited to human behaviour but which encompasses computational models of vocal interactivity in and between humans, animals and robots. It is my view that progress in this interdisciplinary area will unlock key behaviours for interactive systems, and will pave the way for much more effective human-machine interfaces – especially if they involve spoken language.

I've managed to conduct some preliminary research in the area. For example, I've performed experiments using e-Puck robots interacting and vocalising using a novel general-purpose mammalian vocal synthesiser configured to generate rat-squeaks (see <http://www.youtube.com/watch?v=E4aMHK7AH5M>). Likewise, I've been working with Zeno (the

RoboKind humanoid robot) to investigate synchronous agent-to-agent behaviour within a PRESENCE framework. I've also been modelling the consequences of category misalignment between different modes of interactivity (visual, vocal and behavioural), which led to my 2013 Nature paper presenting the first quantitative model of the well-known 'uncanny valley' effect.

My overall aim is to demonstrate that many of the little-understood paralinguistic features exhibited in human speech (including prosody and emotion) are derived from characteristics that are shared by living systems in general. Modelling such behaviours in this wider (situated and embodied) context, using robots as an experimental platform, should eventually enable us to implement usable and effective interaction between human beings and artificial intentional agents. The research aims to address fundamental interactive behaviours such as mimicry, imitation, adaptation, learning, speaker-listener coupling, acquisition, evolution and cooperative/competitive social interaction.

References

- 1 Moore, R. K. *Spoken language processing: piecing together the puzzle*. *Speech Communication*, 49(5), 418–435, 2007
- 2 Moore, R. K. *PRESENCE: A human-inspired architecture for speech-based human-machine interaction*. *IEEE Trans. Computers*, 56(9), 1176–1188, 2007
- 3 Moore, R. K. *A Bayesian explanation of the “Uncanny Valley” effect and related psychological phenomena*. *Nature Scientific Reports*, 2(864), doi:10.1038/srep00864, 2012

5.14 Statement by Julie Oswald

Julie Oswald (University of St Andrews, GB)

License © Creative Commons BY 3.0 Unported license
© Julie Oswald

Acoustic species recognition in delphinids

Dolphin species tend to be acoustically active and produce a variety of sounds. However, many acoustic recordings of dolphins do not have associated visual observations and it can be difficult to identify species in the recordings. Whistles produced by dolphins are narrowband, tonal sounds that are believed to function as social signals and carry information related to individual identity, arousal state and possibly other information such as species identity. As such, much research in recent years has focused on developing tools for classifying whistles to species. Whistle contour shapes are highly variable within species and exhibit a great deal of overlap in time-frequency characteristics when compared between species, which makes classification of these sounds challenging. There are several facets of this topic that are especially relevant in a comparative VIHAR context:

- Computational methods for classification: Many statistical and machine learning techniques have been employed, with varying levels of success, to create classifiers, including random forest analysis, Hidden Markov Models, clustering algorithms, neural networks, and others. Collaborations between computer scientists, bioacousticians, signal processors, etc. are crucial for developing the best classifiers.
- Big data: Large datasets that encompass the variability in vocal repertoires are necessary for training successful classifiers. These data can be difficult to compile, organize, share and store. What are the best practices for dealing with these big datasets?

- Between-species communication: Acoustic species identification is important for researchers, but how important is it for dolphins? Do dolphins use whistles to communicate species identity? If so, what features are they attending to for this information? These questions require collaborations among scientists from fields such as cognition, animal behaviour, bioacoustics, signal processing, and others.
- Applications from research on human speech and communication in other taxa: Lessons learned from research on communication in other taxa may give insight to inter-species communication in dolphins and acoustic species recognition in these species.

5.15 Statement by Bhiksha Raj

Bhiksha Raj (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 3.0 Unported license
© Bhiksha Raj

When to interrupt: a comparative analysis of interruption timings within collaborative communication tasks

This study seeks to determine if it is necessary for the software agent to monitor the communication channel to effectively detect appropriate times to convey information or “interrupt” the operator in a collaborative communication task between a human operator and human collaborators. There is empirical research dedicated to manipulating time on the delivery [Bailey and Konstan 2006; Czerwinski et al. 2000b; Monk et al. 2002] of system-mediated interruptions [McCrickard et al. 2003] in multi-task environments [McFarlane and Latorella 2002]. There is also literature that explores immediate interruption or notification dissemination [Czerwinski et al. 2000a; Dabbish and Kraut 2004; Latorella, 1996] within dual-task scenarios. Studies have shown that delivering interruptions at random times can result in a decline in performance on primary tasks [Bailey & Konstan 2006; Czerwinski et al. 2000a; Kreifeldt and McCarthy 1981; Latorella, 1996; Rubinstein et al. 2001]. Additionally, studies have illustrated that interrupting users engaged in tasks has a considerable negative impact on task completion time [Cutrell et al. 2001; Czerwinski et al. 2000a, 2000b; Kreifeldt and McCarthy 1981; McFarlane 1999; and Monk et al. 2002]. Much of the current literature is focused on one user engaged in a primary task interrupted by a peripheral task.

This study differs from previous studies in that the primary task is collaboration between two or more users and the secondary task is presented to one of the collaborating users. This study explores the outcome of overall task performance and time of completion (TOC) of a task at various delivery times of periphery task interruptions. The study attempts to determine if there is a need for a system to monitor a collaborative communication channel prior to disseminating interruptions that improves efficient communication and prevents information overload within a human exchange. The study uses a simulated collaborative, goal-oriented task via a dual-task where an operator participates in the primary collaborative communication task and a secondary monitoring task. User performance at various interruption timings: random, fixed, and human-determined (HD) are evaluated to determine whether an intelligent form of interrupting users is less disruptive and benefits users’ overall interaction.

There is a significant difference in task performance when HD interruptions are delivered in comparison with random and fixed timed interruption. There is a 54% overall accuracy for task performance using HD interruptions compared to 33% for fixed interruptions and

38% for random interruptions. Additionally when the TOC for the dual-task is compared across interruption types, the TOC for HD interruptions is lower than fixed and randomly timed interruptions. Although on average users complete the dual-task in less time when the communication channel is monitored, the TOC averages are close and there is no significant difference in the completion times. Results show that the use of HD interruptions results in improved task performance in comparison to fixed and randomly timed interruptions. These results are promising and provide some indication that monitoring a communication channel or adding intelligence to the interaction can be useful for the exchange.

5.16 Statement by Rita Singh

Rita Singh (Carnegie Mellon University – Pittsburgh, US)

License  Creative Commons BY 3.0 Unported license
© Rita Singh

Thoughts on human-human, human-animal and human-robot interactions

Currently, at the time of writing this report, shortly after the completion of VIHAR at Schloss Dagstuhl, my primary focus is on the application of Artificial intelligence techniques to voice forensics. Specifically, I work on profiling humans from their voice. Profiling in this context refers to the generation of a complete description of the speaker's persona from their voices. This includes the deduction of the physical appearance, medical status, demographic, sociological and other parameters of a person, and also the person's surroundings, entirely from their voice. In my recent work with the US Coast Guard Investigative Services, I have analyzed scores of hoax distress calls transmitted over national distress channels, and have provided physical descriptions of the perpetrators, of their location and their equipment sufficiently accurately to enable significant success in the investigative process. The ability to track and describe humans through their voice is useful in several disciplines of intelligence, where voice is part of the intelligence information gathered.

The relevance of my work to VIHAR is founded on the methodology I use for this work. My work builds on the fact that humans make numerous judgments about other people from their voices, such as their gender, emotional state, state of health, intelligence etc. There have been hundreds of studies on the ability of humans to make a surprisingly diverse range of judgments about other people entirely from their voices. My approach involves the utilization of AI techniques to achieve super-human capabilities that enable machines to make faster, more accurate, more abundant and deeper assessments of people from their voices. The methodology that I have developed for this is called micro-articulometry. It involves using state-of-art automatic speech recognition and audio processing technologies, to fragment voice recordings into pattern-consistent segments of very short durations, with high precision in high-noise environments. Scores of different "micro-features" are then extracted from these (processed) fragments. These are characteristics of the signal that are usually not observable or measurable by humans manually, and carry signatures of the speaker's persona, upbringing, medical conditions etc. in a manner similar in concept to DNA-biomarker encoding. Amongst other things, the list includes signatures of the physical environment in which the voice was produced and the devices and mechanisms that were used to transmit it. This derived information then feeds into relevant AI techniques designed for learning and discovery from ensembles of data. Suitable machine learning algorithms are then used to "derive" or "predict" the speaker's persona from these micro features. I hope to be able to build physically accurate holograms of humans from their voices in the future.

This methodology has direct relevance to the goals of VIHAR: especially that of enhancing the effectiveness of interactions. Auditory judgments are an aspect of human (and perhaps animal) sensory intelligence that have not been tapped into as a resource for interaction-enhancement until recently. In my interactions with the diverse community of human-human, human-robot and human-animal interactivity researchers at VIHAR, I explored the viability of using the micro-characteristics of human voice – which I also refer to as infra-sensory information, since it is often neither under voluntary control of the speaker, nor consciously perceivable by the human – to enable both robots and animals to understand humans better.

VIHAR was a tremendously enriching experience for me in some ways. From my colleagues who work on human-robot interactions, I learned about their techniques for simulating emotional intelligence in robots. I was able to easily see how my approaches in forensics could enhance those simulations by allowing for subtle reactions in robots in response to changes in the voice (and by association the physical and mental status) of the humans they interact with. My colleagues who work on the analysis and understanding of animal vocalizations in different settings, and on human-animal interactions changed my perspective of the field of interactivity in general. I now believe that while animals may not have the capability of understanding human language as humans do, they may nevertheless be able to discern changes in voice (and speech) patterns at multiple levels, and may be taught to react appropriately to them. I was able to make this hypothesis after listening to presentations, and participating in discussions about animal vocalizations with my colleagues. The vice-versa may also be possible, where humans may be able to interpret the nuances in animal behavior more meaningfully by utilizing the computational techniques that I now use in forensic profiling to enhance their ability to interpret animal sounds. I now firmly believe that AI systems may be able to revolutionize the field of human-animal interaction.

5.17 Statement by Dan Stowell

Dan Stowell (Queen Mary University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Dan Stowell

How do we model vocalisations in general, across hundreds of species?

Speech research has had the luxury of focusing on a single species, tailoring models for that one species' communication system. Furthermore the models developed are often tailored to a single task (e.g. speech recognition vs speaker identification). Many animal communications researchers also focus on specific species or taxa, concentrating on the aspects that are particularly salient for their questions.

We wish to model vocalisations for general-purpose multi-species machine listening. To do this – especially for sequences of vocalisations (whether intra- or inter-individual) – we need models that can capture enough of the relevant details, yet which are generic enough to be reused across widely different species.

A classic approach has been to transliterate animal sounds as sequences of symbols (ABBBABBBABABC), and to study the resulting sequences using n-grams or Markovian models. The reader might miss the discretisation issues swept under the carpet: how was the continuous sound stream segmented into units, and how were those units assigned one of a small set of labels? In a few studies, correspondence with categorical perception in the target species has been measured, but more commonly we trust a human listener or a

clustering algorithm. Even where the discretisation does match up with perceptual categories, it obscures qualitative modulation (how it was said) and the fact that one signal can carry multiple meanings simultaneously. It also usually discards all timing information, while it is clear that the timing of the intervals between units is structured (even if not meaningful) in many species.

In my own work with bird sounds I have recently focused on models which properly integrate the timing of vocalisations. These are to be integrated with analyses of the content of vocal units. Even within the songbirds we have a massive variety of communication styles (tonal vs. non-tonal; simple vs. richly-structured; solos, duets).

Various basic paradigms are available. Markovian models such as HMMs and their extensions. Point processes. Deep learning such as RNNs. Is any of these paradigms appropriate, sufficient?

Modelling ‘state’ in a sound scene: how little can we get away with?

If we are to develop machines that can make sense of vocalisations in multi-species sound scenes, we need models that can reflect all the important aspects of a sound scene, which presumably includes some information relating to the actors in the scene. Yet the true ‘internal state’ of those actors may be quite different depending on their species – birds, humans, animals, robots – and we do not want to be forced to specify a highly-complex model for every species we might encounter. Do we wish to maintain a model of each ‘agent’ detected in a sound scene – or can we get away without it, leaving it implicit in the network of individual vocalisations affecting each other and the observer?

Assuming that we do want a model of agents, is there a minimal ‘internal state’ model that can be applied in many situations? Theory-of-mind approaches tend to imply a rich model of agents with beliefs, motivations, affordances. At the opposite end, a basic HMM-like model of actors could contain nothing more than a small set of unlabelled states. Is there something useful between these, e.g. information access combined with the handy circumplex model of affect?

Not all state is captured in the agents: contextual variables come in too, such as the temperature or the noise background.

Machine audition – how can it ever be as robust and flexible as human/animal audition?

However impressive recent results have been, there remains quite a gulf between the performance of any given machine and human/animal audition. Whether through inherited or learnt structure, we are able to cope with a vast array of: interfering noises (weather, traffic, television); modifications that the environment makes to a sound (reflections in a forest, reflections from a wall, frequencydependent attenuation, turbulence); novel sounds.

Can such robustness and flexibility be represented in computational methods economically? (e.g. without having to model general learning)

Reaching across disciplines

A practical challenge: relevant disciplines here include bioacoustics, ecology, animal behaviour, signal processing, machine learning, robotics, etc. – and to address these issues we need sustained cross-disciplinary engagement. The various disciplines often have different ideas about what can be taken for granted, which conferences to go to – and how much a conference should cost...

5.18 Statement by Zheng-Hua Tan

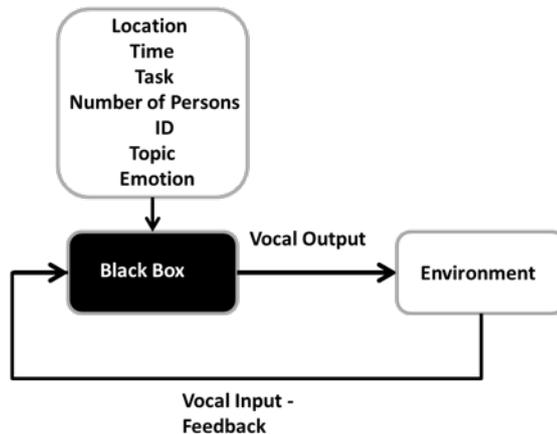
Zheng-Hua Tan (Aalborg University, DK)

License  Creative Commons BY 3.0 Unported license
© Zheng-Hua Tan

Durable vocal interactive system for socially intelligent robots

A social robot should be able to interact in a meaningful way with its users and to maintain a long-term relationship with them. To meet this requirement, a durable vocal interactive system is very important, as speech interaction is perhaps the most important aspect of interpersonal communication. A vocal interactive system for robots should not only operate as a simple answer machine, but should also understand and respond accordingly to different users and different situations. For example, the robot should have memory of its users and past conversations with them in order to be able to sustain a long-term interaction. Furthermore, it should extract the scene information and also the information of the users during the conversation in order to have natural vocal interaction just like in human communication.

We aim to develop a durable vocal interactive system as described in the figure below, where we have a black box (machine learning methods) which can utilize the conversation information, take the vocal input and sensor the environment as feedback in order to give natural vocal output. This black box is capable of life-long learning, which means it can model its users, extract and remember information from their past conversations and selfadapt based on this information.



To develop such a system, several challenges arise:

1. We need to figure out how to model the vocal interaction between humans and robots, which can take all the related input into account to generate the output and also save and manage extracted information.
2. What input aspects are important for human robot interaction (HRI)? E.g., location, time, task, number of persons, identity of persons, topic of discussion, communicated emotion and etc.
3. What kind of explicit/implicit methods to use for acquiring the input aspects? Similar to reinforcement learning, could we fuse the input aspects using reinforcement fusion?
4. Can reinforcement learning be applied to a vocal interaction system? If the answer is yes, what language, or code phrases, or key terms should be used as a reward (positive/negative feedback) for the robotic system if we use reinforcement learning?

5. How to realise life-long learning for a vocal interactive social robot, which can enable the robot being self-adapting to the environment and users?

Project iSocioBot (Durable Interaction with Socially Intelligent Robots):
<http://socialrobot.dk/>

5.19 Statement by Serge Thill

Serge Thill (University of Skövde, SE)

License  Creative Commons BY 3.0 Unported license
 © Serge Thill

Concept grounding has received much attention over the past few decades. It is arguably one of the defining aspects of embodied cognitive science as a breach with good old-fashioned computationalism (as opposed to embodied cognitive science as a continuation of American naturalism and ecological psychology as championed, for example, by Chemero). The exact degree, if any, to which a concept needs to be grounded in an agent’s own experience however remains a matter of much debate. On one hand, there is plenty of evidence for the involvement of sensorimotor cortices in the processing of language (particularly words that directly relate to the sensorimotor aspect in question, see [5], for a discussion). On the other hand, one cannot deny that computational linguistics has made significant progress in machine understanding of language over the past decades, based on statistical information alone [4].

My interest is in trying to characterise what concepts are “made up from” internally. This can include direct grounding in sensorimotor experience (where I defend the idea that sensorimotor experience must be considered to encompass more than just interaction with the external world – interoception matters just as much and it is not sufficient to merely consider the perception of a given sound attached to some other (e.g. visual) input), but also statistical information, and information conveyed by others (via, for example metaphors in the Lakoff & Johnson sense). An initial stab at formulating such a characterisation – using Chris Eliasmith’s *semantic pointer architecture* for a number of reasons [3] – is presented in [1].

My main reason for investigating such questions is because I am interested in social artificial cognitive systems, and what it takes to create some that are natural to interact with. In this context, there are two points that need to be made. The first is that artificial agents need to be able to understand human concepts (rather than the other way around) to be intuitive to interact with. The second is that, if theories of embodiment are right, then the inevitable differences between robot and human bodies are bound to limit this understanding since there will be limitations to the degree to which a human concept can be grounded in a robotic body (this is particularly true when we consider interoceptive aspects of concept grounding [1]).

To create social artificial agents, we must therefore find a way to overcome these differences, which implies that we need to understand how, precisely, the body may matter in concept formation. It is here that studying vocal interaction between all types of living beings – with all the commonalities and variation in embodied and sensorimotor experience this implies – as well as explicitly trying to build machines that can offer the same types of interaction despite having a different embodied experience of the concepts used will help to further the state of the art. Of particular interest from the perspective of creating social robots are the possibilities and limitations in interaction between non-conspecifics [4].

All that said, there is still much for me to learn about animal vocalisations. While it seems relatively uncontroversial to me to state that studying vocalisations across a wide range of embodied experiences will shed light onto the role that such an experience plays, the exact approach by which this potential can be unlocked remains to be explored in the context of VIHAR. Similarly, all of the above conflates vocalisation and meaningful communication. While I won't attempt to separate these here in the hope to retain a somewhat succinct statement (but see [2], for a somewhat more detailed discussion), it is very clear that this is going to substantially shape research in this direction.

References

- 1 Thill S., Twomey K. *What's on the inside counts: A grounded account of concept acquisition and development* in *Frontiers in Psychology: Cognition*, 7(402), 2016.
- 2 Moore R. K., Marxer R., Thill S. *Vocal interactivity in-and-between humans, animals and robots* in *Frontiers in Robotics and AI*, 3(61), 2016.
- 3 Thill S. *Embodied neuro-cognitive integration* in *Proceedings of the Workshop on "Neural-Cognitive Integration" (NCI@KI 2015)*, 2015.
- 4 Thill S., Padó S., Ziemke T. *On the importance of a rich embodiment in the grounding of concepts: perspectives from embodied cognitive science and computational linguistics* in *Topics in Cognitive Science*, 6(3), p. 545–558, 2014.
- 5 Chersi F., Thill S., Ziemke T., Borghi A. M. *Sentence processing: linking language to motor chains* in *Frontiers in Neurorobotics*, 4(4), 2010.

5.20 Statement by Petra Wagner

Petra Wagner (Universität Bielefeld, DE)

License  Creative Commons BY 3.0 Unported license
© Petra Wagner

What do we minimally need to communicate and how do we assess that communication is working?

As the concept of language is meaningless without assuming its being shared and used interactively by a linguistic community (Wittgenstein's private language argument), the investigation of communicative vocalizations does not make sense from a solipsistic, monadic perspective. Unfortunately, much work in linguistics has done exactly this and has focused on aspects of either production, perception or grammatical intuition. Among many other things, we are therefore surprisingly ignorant about the prerequisites of felicitous interactions. This deficit makes it all the more difficult to determine which aspects of linguistic or proto-linguistic communicative skills should be necessarily realized in artificial systems.

In any new communicative encounter with con-species (e.g. speaking a different language) or other interlocutors (e.g. artificial systems, animals, aliens), we need to negotiate how a process of informational grounding can be successfully implemented. To achieve this, we seem to rely on an a priori set of communicative "customs" which ultimately pave the way for communicative interaction, or, rather a common ground concerning how communication works. This common ground needs to be explored further and HRI provides a very useful platform for this endeavor.

I hypothesize that at least from a human perspective, this a priori common ground would have to include the following:

1. Some fundamental mechanism(s) organizing the sequentiality or simultaneity of interlocutors' vocalizations, e.g. by anticipating upcoming speech onsets and terminations (e.g. by respiratory cues, slowing down, falling intonation) together with some general customs for organizing the floor exchange.
2. The expression of general responsiveness and some agreement on what signals this responsiveness (feedback, attention, entrainment?)
3. An initially very coarse pool of shared signs, e.g. related to universal concepts (e.g. the frequency code, where f_0 expresses size or movement direction or the effort code, where more articulatory effort expresses relevance, e.g. danger or surprise).
4. The presupposition that shared signs are flexible in the sense that they can be transferred into other system architectures (e.g. those equipped with different sound production mechanisms) or modalities (e.g. gestures).

We have worked to some degree on all these issues, trying to understand better the communicative function of subtle phonetic cues such as inhalations [2], disfluencies in (synthetic) speech [1], the success of an entrainment-based multimodal feedback mechanism in an artificial agent [3], the multimodal expression of attention [4], the usage of iconic prosody and speech-movement synchronization in a coaching scenario [5] and the flexibility of cues, extending to the domain of co-speech gestures [7, 5]. While we are still only beginning to comprehend these various complex communicative factors, we furthermore believe that we need to find novel methodological paradigms to investigate interactions both “in the wild” and under more controlled laboratory conditions [6].

Naturally, I believe that most of my assumptions sketched above are false or at least need a lot more thinking, extension and exploration. But in order to assess the general applicability of these ideas, we need to come up with working methodological paradigms on how to properly assess whether an interaction is perceived as felicitous by the interlocutors. Unfortunately, in my opinion, we currently lack suitable online (!) approaches to evaluate HRI or ongoing human-human interactions.

References

- 1 Betz, S., Wagner, P., & Schlangen, D. *Micro-Structure of Disfluencies: Basics for Conversational Speech Synthesis*. Proceedings of Interspeech 2015, Dresden, 2015.
- 2 Cwiek, A., Neueder, S., & Wagner, P. *Investigating the communicative function of breathing and non-breathing “silent” pauses*. PundP 12 – Phonetik und Phonologie im deutschsprachigen Raum. München, 2016.
- 3 Inden, B., Malisz, Z., Wagner, P., & Wachsmuth, I. *Timing and entrainment of multimodal backchanneling behavior for an embodied conversational agent*. In J. Epps, F. Chen, S. Oviatt, K. Mase, A. Sears, K. Jokinen, & B. Schuller (Eds.), Proceedings of the 15th International Conference on Multimodal Interaction, ICMI'13 - Sydney New York: ACM, 2013.
- 4 Malisz, Z., Włodarczak, M., Buschmeier, H., Skubisz, J., Kopp, S., & Wagner, P. *The ALICO Corpus: Analysing the Active Listener*. Language Resources and Evaluation, 50(2), 2016.
- 5 Skutella, L. V., Süßenbach, L., Pitsch, K., & Wagner, P. *The prosody of motivation. First results from an indoor cycling scenario*. In R. Hoffmann (Ed.), Studentexte zur Sprachkommunikation: Vol. 71. Elektronische Sprachsignalverarbeitung 2014 (pp. 209–215). TUD Press, 2014.
- 6 Wagner, P., Trouvain, J., & Zimmerer, F. *In defense of stylistic diversity in speech research*. Journal of Phonetics, 48, 1–12, 2015.
- 7 Wagner, P., Malisz, Z., & Kopp, S. *Gesture and Speech in Interaction: An Overview*. Speech Communication, 57(Special Iss.), 209–232, 2014.

5.21 Statement by Benjamin Weiss

Benjamin Weiss (TU Berlin, DE)

License  Creative Commons BY 3.0 Unported license
© Benjamin Weiss

Interpersonal Perception and Evaluation

The first (acoustic) impression results in immediate person attributions. However, which impression is dominant in the listener is still hard to grasp, as it depends on individual expectations and preferences. Several acoustic correlates of such person attributions have already been identified, but we still lack a model of relevant general (physiologically grounded) and individual, i.e. interpersonal, attributions that incorporates non-linear relationships with acoustic and/or articulatory features as well as listener properties (e.g. personality, background, voice). Here, insights from animal vocalizations might be fruitful to consider.

A subsequent evaluation of the dialog partner (e.g. on competence/benevolence and likeability), or even of the dialog (e.g. satisfaction), can only be successfully studied, if the most salient attributions of a dialog partner are known, and the situational context is taken into account. This challenge is even higher when moving from simple listening situations to interactive ones, as the attributions and attitudes towards the speaker will be reflected in conversational behavior.

When applying results from HHI to HRI, e.g. backchannel or turn-taking signals, there arise several methodological issues during evaluation.

- Whereas human observed behavior can mostly be considered as situationally adequate and congruent on multiple linguistic levels (semantic, pragmatic, para-linguistic, nonverbal), current implementations can, of course, only consider one or a few of such levels and features, which might limit validity of evaluations results. Trying to identify different communication strategies might help to reduce complexity without simplifying communication behavior too much (e.g. should a certain degree of acoustic-prosodic entrainment not also be reflected on other linguistic and gestural levels and on back-channeling behavior in a robot?).
- However, even basic communication signals in robots have been found to affect humans in real social situations. Such approaches (even popular in design and arts) seem to be very promising to study social aspects in HRI, or even HHI, than trying to build complex AI. In order to better understand and interpret such results, new methods to assess relevant interaction events are necessary, with the aim to address relevance of vocal signals outside the laboratory.
- Does making a robot more human-like by implementing vocal communication skills really results in a better user experience? The limits of such aims have not yet been explored to a satisfying degree.

References

- 1 Weiss, B., F. Burkhardt & M. Geier *Towards perceptual dimensions of speakers' voices: Eliciting individual descriptions*. Proc. Workshop on Affective Social Speech Signals, Grenoble, 2013.
- 2 Weiss, B. & K.Schoenberg *Conversational structures affecting auditory likeability*. Proc. Interspeech, 1791–1795l, 2014.
- 3 Weiss, B. & S Hillmann *Feedback Matters: Applying Dialog Act Annotation to Study Social Attractiveness in Three-Party Conversations*. 12. Joint ACL – ISO Workshop on Interoperable Semantic Annotation, Portorož, pp. 55–58, 2016.

- 4 Fernandez Gallardo, L. & B. Weiss *Speech Likability and Personality-based Social Relations: A Round-Robin Analysis over Communication Channels*. Proc. Interspeech. pp. 903–907, 2016.

Participants

- Andrey Anikin
Lund University, SE
- Timo Baumann
Universität Hamburg, DE
- Tony Belpaeme
University of Plymouth, GB
- Elodie Briefer
ETH Zürich, CH
- Nick Campbell
Trinity College Dublin, IE
- Fred Cummins
University College Dublin, IE
- Angela Dassow
Carthage College – Kenosha, US
- Robert Eklund
Linköping University, SE
- Julie E. Elie
University of California –
Berkeley, US
- Sabrina Engesser
Universität Zürich, CH
- Sarah Hawkins
University of Cambridge, GB
- Ricard Marxer
University of Sheffield, GB
- Roger K. Moore
University of Sheffield, GB
- Julie Oswald
University of St Andrews, GB
- Bhiksha Raj
Carnegie Mellon University –
Pittsburgh, US
- Rita Singh
Carnegie Mellon University –
Pittsburgh, US
- Dan Stowell
Queen Mary University of
London, GB
- Zheng-Hua Tan
Aalborg University, DK
- Serge Thill
University of Skövde, SE
- Petra Wagner
Universität Bielefeld, DE
- Benjamin Weiss
TU Berlin, DE

