



Volume 5, Issue 5, May 2015

Compositional Verification Methods for Next-Generation Concurrency (Dagstuhl Seminar 15191) <i>Lars Birkedal, Derek Dreyer, Philippa Gardner, and Zhong Shao</i>	1
The Message in the Shadow: Noise or Knowledge? (Dagstuhl Seminar 15192) <i>Roberto Casati, Patrick Cavanagh, and Paulo E. Santos</i>	24
Cross-Lingual Cross-Media Content Linking: Annotations and Joint Representations (Dagstuhl Seminar 15201) <i>Alexander G. Hauptmann, James Hodson, Juanzi Li, Nicu Sebe, and Achim Rettinger</i>	43
Theory of Evolutionary Algorithms (Dagstuhl Seminar 15211) <i>Benjamin Doerr, Nikolaus Hansen, Christian Igel, and Lothar Thiele</i>	57
Multi-disciplinary Approaches to Reasoning with Imperfect Information and Knowledge – a Synthesis and a Roadmap of Challenges (Dagstuhl Seminar 15221) <i>Igor Douven, Gabriele Kern-Isberner, Markus Knauff, and Henri Prade</i>	92
Human-Centric Development of Software Tools (Dagstuhl Seminar 15222) <i>Andrew J. Ko, Shriram Krishnamurthi, Gail Murphy, and Janet Siegmund</i>	115

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

January, 2016

Bibliographic information published by the Deutsche Nationalbibliothek

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

License

This work is licensed under a Creative Commons Attribution 3.0 DE license (CC BY 3.0 DE).



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

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

The copyright is retained by the corresponding authors.

Aims and Scope

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

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

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Editorial Board

- 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
- Michael Waidner
- Reinhard Wilhelm

Editorial Office

Marc Herbstritt (*Managing Editor*)
Jutka Gasiorowski (*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.5.5.i

Compositional Verification Methods for Next-Generation Concurrency

Edited by

Lars Birkedal¹, Derek Dreyer², Philippa Gardner³, and Zhong Shao⁴

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

2 MPI-SWS – Saarbrücken, DE, dreyer@mpi-sws.mpg.de

3 Imperial College London, GB, pg@doc.ic.ac.uk

4 Yale University, US, zhong.shao@yale.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15191 “Compositional Verification Methods for Next-Generation Concurrency”. The seminar was successful and facilitated a stimulating interchange between the theory and practice of concurrent programming, and thereby laid the ground for the development of compositional verification methods that can scale to handle the realities of next-generation concurrency.

Seminar May 3–8, 2015 – <http://www.dagstuhl.de/15191>

1998 ACM Subject Classification D1.3 Concurrent Programming, D.2.4 Software/Program Verification, D.3 Programming Languages

Keywords and phrases Verification of Concurrent Programs (Models, Logics, Automated Analysis), Concurrent Programming

Digital Object Identifier 10.4230/DagRep.5.5.1

1 Executive Summary

Lars Birkedal

Derek Dreyer

Philippa Gardner

Zhong Shao

License © Creative Commons BY 3.0 Unported license
© Lars Birkedal, Derek Dreyer, Philippa Gardner, and Zhong Shao

One of the major open problems confronting software developers today is how to cope with the complexity of reasoning about large-scale concurrent programs. Such programs are increasingly important as a means of taking advantage of parallelism in modern architectures. However, they also frequently depend on subtle invariants governing the use of shared mutable data structures, which must take into account the potential interference between different threads accessing the state simultaneously. Just figuring out how to express such invariants at all has proven to be a very challenging problem; even more challenging is how to support *local* reasoning about such invariants, i.e., confining the reasoning about them to only the components of the program that absolutely need to know about them.

Fortunately, we are now at a point where verification research has produced the critical foundations needed to tackle this problem: namely, *compositional* methods, which exploit the inherently modular structure of realistic concurrent programs in order to decompose



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

Compositional Verification Methods for Next-Generation Concurrency, *Dagstuhl Reports*, Vol. 5, Issue 5, pp. 1–23
Editors: Lars Birkedal, Derek Dreyer, Philippa Gardner, and Zhong Shao



Dagstuhl Reports

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

verification effort along module boundaries. Fascinatingly, a variety of different but related compositional methods have been developed contemporaneously in the last several years:

- **Separation logics:** Separation logic was developed initially as a generalization of Hoare logic – supporting local, compositional reasoning about sequential, heap-manipulating programs – and much of the early work on separation logic has been successfully incorporated into automated verification tools like Smallfoot [2], SLAyer [3], Abductor [6], etc., scaling to handle millions of lines of code. Recently, there have been a series of breakthroughs in adapting separation logic to handle concurrent programs as well. Concurrent separation logic [17] provides course-grained local reasoning about concurrent programs; combining this local reasoning with rely-guarantee reasoning [26] provides fine-grained concurrent reasoning; intertwining abstraction with local reasoning enables a client to reason about the use of a set module [8] without having to think about the underlying implementation using lists or concurrent B-trees; and, very recently, all this has been extended to account for higher-order programs as well [21].
- **Kripke models:** There is a long line of work on the use of semantic models like *Kripke logical relations* [1, 9] (and more recently *bisimulations* [19, 20]) for proving observational equivalence of programs that manipulate local state. Observational equivalence is useful not only for establishing correctness of program transformations (e.g., in compiler certification) but also as a verification method in its own right (e.g., one can prove that a complex but efficient implementation of an ADT is equivalent to a simple but inefficient reference implementation). However, it is only in the last few years that such models have been generalized to account for the full panoply of features available in modern languages: higher-order state, recursion, abstract types, control operators, and most recently concurrency, resulting in some of the first formal proofs of correctness of sophisticated fine-grained concurrent algorithms in a higher-order setting [1, 9, 23]. These advances have come about thanks to the development of more elaborate Kripke structures for representing invariants on local state.
- **Hoare type theory:** Dependent type theory provides a very expressive compositional verification system for higher-order functional programs, so expressive that types can characterize full functional correctness. Traditionally, however, dependent type theories were limited to verification of *pure* programs. Recent work on Hoare type theory (HTT) [15] has shown how to integrate effects into dependent type theory by incorporating Hoare triples as a new primitive type, and prototypes of HTT have been implemented in Coq [7, 16], allowing for imperative programs to be verified mechanically as they are being written. Moreover, first steps of extending HTT with concurrency have recently been taken [14], thus giving hope for a potential future integration of design and verification for higher-order concurrent programs.

All in all, the field of modular concurrency verification is highly active, with groundbreaking new developments in these and other approaches coming out every year. Particularly fascinating is the appearance of deep connections between the different methods. There are striking similarities, for instance, between the advanced Kripke structures used in recent relational models of higher-order state and the semantic models underlying recent concurrent separation logics.

Nevertheless, there are a number of ways in which the advanced models and logics developed thus far are still, to be honest, in their infancy. Most of these approaches, for example, have only been applied to the verification of small, self-contained ADTs and have not yet been scaled up to verify large-scale modular concurrent programs. Moreover, even

the most state-of-the-art compositional methods do not yet account for a number of the essential complexities of concurrent programming as it is practiced today, including:

- **Weak memory models:** The vast majority of state-of-the-art compositional verification methods are proved sound with respect to an operational semantics that assumes a sequentially consistent memory model. However, modern hardware implements weak memory models that allow for many more reorderings of basic operations. Thus there is a clear gap between the verification theory and practice that needs to be filled (for efficiency reasons we, of course, do not want to force programmers/compiler to insert enough memory fence operations to make the hardware behave sequentially consistent). This problem has been known for the last decade, but it is only in the last year or two that formal descriptions of the behavior of programming languages with weak memory models have been developed. Given this foundation, we should now be able to make progress on extending compositional verification methods to weak memory models.
- **Higher-order concurrency:** Higher-order functional abstraction is an indispensable feature of most modern, high-level programming languages. It is also central to a variety of concurrent programming idioms, both established and nascent: work stealing [4], Concurrent ML-style events [18], concurrent iterators [13], parallel evaluation strategies [22], STM [11], reagents [24], and more. Yet, only a few existing logics have been proposed that even attempt to account for higher-order concurrency [21, 14, 12], and these logics are just first steps – for example, they do not presently account for sophisticated “fine-grained” concurrent ADTs. Verification of higher-order concurrent programs remains a largely open problem.
- **Generalizing linearizability:** Sophisticated concurrent data structures often use fine-grained synchronization to maximize the possibilities for parallel access. The classical correctness criterion for such fine-grained data structures is *linearizability*, which ensures that every operation has a linearization point at which it appears (to clients) to atomically take effect. However, existing logics do not provide a way to exploit linearizability directly in client-side reasoning, and moreover the notion does not scale naturally to account for operations (such as higher-order iterators) whose behavior is not semantically atomic. Recently, researchers have started to investigate alternative approaches, based on *contextual refinement* [10, 23]. And methods for reasoning about operations with multiple linearizability points are also being developed.
- **Liveness properties:** Synchronization of concurrent data structures can also affect the progress of the execution of the client threads. Various progress properties have been proposed for concurrent objects. The most important ones are wait-freedom, lock-freedom and obstruction-freedom for non-blocking implementations, and starvation-freedom and deadlock-freedom for lock-based implementations. These properties describe conditions under which method calls are guaranteed to successfully complete in an execution. Traditional definitions (which are quite informal) of these progress properties are difficult to use in modular program verification because they fail to describe how the progress properties affect clients. It is also unclear how existing separation logics, which were primarily designed for proving partial correctness, can be adapted to prove progress properties. Recently, researchers have started to combine *quantitative reasoning* of resource bounds with separation logics, which offer new possibilities for verifying both safety and liveness properties in a single framework.

Grappling with these kinds of limitations is essential if our verification technology is to be relevant to real-world programs running on modern architectures, and as such it poses exciting new research questions that we as a community are just beginning to explore.

In this seminar, we brought together a wide variety of researchers on concurrency verification, as well as leading experts on concurrent software development in both high- and low-level languages. The goal was to facilitate a stimulating interchange between the theory and practice of concurrent programming, and thereby foster the development of compositional verification methods that can scale to handle the realities of next-generation concurrency.

Among the concrete research challenges investigated in depth during the seminar are the following:

- What are good ways of reasoning about weak memory models? It should be possible to reason about low-level programs that exploit weak memory models (e.g., locks used inside operating systems) but also to reason at higher levels of abstractions for programs that use sufficient locking.
- What is the best way to define a language-level memory model that is nevertheless efficiently implementable on modern hardware. C11 is the state of the art, but it is flawed in various ways, and we heard about a number of different ways of possibly fixing it.
- What is the best way to mechanize full formal verification of concurrent programs, using interactive proof assistants, such as Coq.
- How can we adapt existing and develop new compositional techniques for reasoning about liveness properties of concurrent programs? Can we apply quantitative techniques to reduce the proof of a liveness property to the proof of a stronger safety property? Also, recent work on rely-guarantee-based simulation can prove linearizability of a sophisticated concurrent object by showing the concurrent implementation is a contextual refinement of its sequential specification. We would hope that similar techniques can be used to prove progress properties as well.
- Only recently have researchers begun to propose logics and models for higher-order concurrency [23, 21]. What are the right concurrency abstractions for higher-order concurrent programming idioms as diverse as transactional memory [11], Concurrent ML [18], joins [25], and reagents [24], among others? What is the best way to even specify, let alone verify, programs written in these idioms, and are there unifying principles that would apply to multiple different idioms?
- Most verification work so far has focused on shared-memory concurrency, with little attention paid to message-passing concurrency (except for some recent work on verifying the C# joins library). Can the models and logics developed for the former be carried over usefully to the latter, and what is the connection (if any) with recent work on proof-theoretic accounts of session types [5]? Can session types help to simplify reasoning about some classes of concurrent programs, e.g., those that only involve some forms of message passing and not full shared memory?
- A number of recent Kripke models and separation logics have employed *protocols* of various forms to describe the invariants about how the semantic state of a concurrent ADT can evolve over time. But different approaches model protocols differently, e.g., using enriched forms of state transition systems vs. partial commutative monoids. Is there a canonical way of representing these protocols formally and thus better understanding the relationship between different proof methods?
- There seem to be tradeoffs between approaches to concurrency verification based on Hoare logic vs. refinement (unary vs. relational reasoning), with the former admitting a wider variety of formal specifications but the latter offering better support for reasoning about atomicity. Consequently, a number of researchers are actively working on trying to combine both styles of reasoning in a unified framework. What is the best way to do this?

- To what extent do we need linearizability to facilitate client-side reasoning? Is it possible in many cases for clients to rely on a much weaker specification? And which ways are there to formalize looser notions, e.g. where there are multiple linearization points?
- Now that we are finally developing logics and models capable of verifying realistic concurrent algorithms, can we abstract away useful proof patterns and automate them? What is needed in order to integrate support for concurrent invariants into automated verification tools like SLayer and Abductor?

These different challenges were discussed through talks and discussions by participants, see the list of talk abstracts below.

References

- 1 Amal Ahmed, Derek Dreyer, and Andreas Rossberg. State-dependent representation independence. In *POPL*, 2009.
- 2 Josh Berdine, Cristiano Calcagno, and Peter W. O’Hearn. Smallfoot: Modular automatic assertion checking with separation logic. In *FMCO*, 2005.
- 3 Josh Berdine, Byron Cook, and Samin Ishtiaq. Slayer: Memory safety for systems-level code. In *CAV*, 2011.
- 4 Robert D. Blumofe, Christopher F. Joerg, Bradley C. Kuszmaul, Charles E. Leiserson, Keith H. Randall, and Yuli Zhou. Cilk: An efficient multithreaded runtime system. *JPDC*, 37(1):55–69, August 1996.
- 5 Luís Caires and Frank Pfenning. Session types as intuitionistic linear propositions. In *CONCUR*, 2010.
- 6 Cristiano Calcagno, Dino Distefano, Peter O’Hearn, and Hongseok Yang. Compositional shape analysis by means of bi-abduction. *Journal of the ACM*, 58(6), 2011.
- 7 Adam Chlipala, Gregory Malecha, Greg Morrisett, Avraham Shinnar, and Ryan Wisnesky. Effective interactive proofs for higher-order imperative programs. In *ICFP*, 2009.
- 8 Thomas Dinsdale-Young, Mike Dodds, Philippa Gardner, Matthew Parkinson, and Viktor Vafeiadis. Concurrent abstract predicates. In *ECOOP*, June 2010.
- 9 Derek Dreyer, Georg Neis, and Lars Birkedal. The impact of higher-order state and control effects on local relational reasoning. *Journal of Functional Programming*, 22(4&5):477–528, August 2012.
- 10 I. Filipović, P.W. O’Hearn, N. Rinetzky, and H. Yang. Abstraction for concurrent objects. In *ESOP*, 2009.
- 11 Tim Harris, Simon Marlow, Simon Peyton-Jones, and Maurice Herlihy. Composable memory transactions. In *PPoPP*, 2005.
- 12 Aquinas Hobor, Andrew W. Appel, and Francesco Zappa Nardelli. Oracle semantics for concurrent separation logic. In *ESOP*, 2008.
- 13 Doug Lea. The `java.util.concurrent.ConcurrentHashMap`.
- 14 Ruy Ley-Wild and Aleksandar Nanevski. Subjective auxiliary state for coarse-grained concurrency. In *POPL*, 2013.
- 15 Aleksandar Nanevski, Greg Morrisett, and Lars Birkedal. Hoare type theory, polymorphism and separation. *JFP*, 18(5-6):865–911, 2008.
- 16 Aleksandar Nanevski, Viktor Vafeiadis, and Josh Berdine. Structuring the verification of heap-manipulating programs. In *POPL*, 2010.
- 17 Peter W. O’Hearn. Resources, concurrency and local reasoning. *Theoretical Computer Science*, 375(1–3):271–307, May 2007.
- 18 John H. Reppy. *Higher-order concurrency*. PhD thesis, Cornell University, 1992.
- 19 Davide Sangiorgi, Naoki Kobayashi, and Eijiro Sumii. Environmental bisimulations for higher-order languages. In *LICS*, 2007.

- 20 Eijiro Sumii. A complete characterization of observational equivalence in polymorphic λ -calculus with general references. In *CSL*, 2009.
- 21 Kasper Svendsen, Lars Birkedal, and Matthew Parkinson. Modular reasoning about separation of concurrent data structures. In *ESOP*, 2013.
- 22 P. W. Trinder, K. Hammond, H.-W. Loidl, and S. L. Peyton Jones. Algorithm + strategy = parallelism. *JFP*, 8(1):23–60, January 1998.
- 23 A Turon, J Thamsborg, A Ahmed, L Birkedal, and D Dreyer. Logical relations for fine-grained concurrency. In *POPL*, 2013.
- 24 Aaron Turon. Reagents: expressing and composing fine-grained concurrency. In *PLDI*, 2012.
- 25 Aaron Turon and Claudio Russo. Scalable join patterns. In *OOPSLA*, 2011.
- 26 Viktor Vafeiadis. *Modular fine-grained concurrency verification*. PhD thesis, University of Cambridge, 2008.

2 Table of Contents

Executive Summary

<i>Lars Birkedal, Derek Dreyer, Philippa Gardner, and Zhong Shao</i>	1
--	---

Overview of Talks

Analysing and Optimising Parallel Snapshot Isolation <i>Andrea Cerone</i>	9
Phantom Monitors: A Simple Foundation for Modular Proofs of Fine-Grained Concurrent Programs <i>Adam Chlipala</i>	9
A Calculus for Relaxed Memory <i>Karl Craty</i>	9
Modular Termination Verification for Non-blocking Concurrency <i>Pedro Da Rocha Pinto</i>	10
Speculation in Higher-Order Separation Logics (and why it's tricky) <i>Thomas Dinsdale-Young</i>	10
Compositional C11 Program Transformation <i>Mike Dodds</i>	10
Static Verification of GPU Kernels <i>Alastair F. Donaldson</i>	11
Making Sense of Rust (Work in Preservation) <i>Derek Dreyer</i>	11
An operational approach to relaxed memory models <i>Xinyu Feng</i>	12
Formally Specifying POSIX File Systems <i>Philippa Gardner</i>	12
An Unsophisticated Higher-Order-ish Logic for Modular Specification and Verification of Total Correctness Properties of Fine-Grained Concurrent Imperative Programs <i>Bart Jacobs</i>	13
Reasoning about possible values in concurrency <i>Cliff B. Jones</i>	13
Iris: Monoids and Invariants as an Orthogonal Basis for Concurrent Reasoning <i>Ralf Jung</i>	13
The Push/Pull Model of Transactions <i>Eric Koskinen</i>	14
Curry-Howard for GUIs via Linear Temporal Classical Linear Logic <i>Neel Krishnaswami</i>	14
Owicki-Gries Reasoning for Weak Memory Models <i>Ori Lahav</i>	15

A Program Logic for Contextual Refinement of Concurrent Objects under Fair Scheduling <i>Hongjin Liang</i>	15
Formal Verification and Linux-Kernel Concurrency <i>Paul McKenney</i>	15
Linearizability: Who Really Needs It? <i>Paul McKenney</i>	16
Some Examples of Kernel-Hacker Informal Correctness Reasoning <i>Paul McKenney</i>	16
Designing a Lock-Free Range Management Algorithm <i>Maged M. Michael</i>	16
Viper – A Verification Infrastructure for Permission based Reasoning <i>Peter Mueller</i>	17
Structures with Intrinsic Sharing, Subjectively <i>Aleksandar Nanevski</i>	17
An operational semantics for C/C++11 concurrency <i>Kyndylan Nienhuis</i>	18
Investigating Weak Memory Performance <i>Scott Owens</i>	18
Polarized Substructural Session Types <i>Frank Pfenning</i>	18
An attempt at fixing C11 concurrency <i>Jean Pichon-Pharabod</i>	19
Automated and Modular Refinement Reasoning for Concurrent Programs <i>Shaz Qadeer</i>	19
CoLoSL: Concurrent Local Subjective Logic <i>Azalea Raad</i>	19
Concurrency-Aware Linearizability <i>Noam Rinetzky</i>	20
Anatomy of mechanized reasoning about fine-grained concurrency <i>Ilya Sergey</i>	20
Using Iris as a meta-language for logical relations <i>Kasper Svendsen</i>	21
Verifying Read-Copy-Update in a Logic for Weak Memory <i>Joseph Tassarotti</i>	21
Software verification under weak memory consistency <i>Viktor Vafeiadis</i>	21
Open Problems and State-of-Art of Session Types <i>Nobuko Yoshida</i>	22
Participants	23

3 Overview of Talks

3.1 Analysing and Optimising Parallel Snapshot Isolation

Andrea Cerone (IMDEA Software – Madrid, ES)

License © Creative Commons BY 3.0 Unported license
© Andrea Cerone

Joint work of Bernardi, Giovanni; Cerone, Andrea; Gotsman, Alexey, Yang; Hongseok

Large-scale Internet services often rely on distributed databases that provide consistency models for transactions weaker than serialisability. Unfortunately, we currently lack a systematic understanding of when programmers can use such models without violating correctness. And when an application is correct on a given consistency model, we do not know whether the model can safely be weakened even further to improve performance.

I will present work in progress to address these issues. In the talk I will concentrate on a promising consistency model of Parallel Snapshot Isolation (PSI), which weakens the classical snapshot isolation in a way that allows more efficient distributed implementations. I will present a formalisation of PSI, a criterion for ensuring correctness of applications using it, and a way of optimising the applications to improving performance.

3.2 Phantom Monitors: A Simple Foundation for Modular Proofs of Fine-Grained Concurrent Programs

Adam Chlipala (MIT – Cambridge, US)

License © Creative Commons BY 3.0 Unported license
© Adam Chlipala

Joint work of Bell, Christian J.; Lesani, Mohsen; Malecha, Gregory; Boyer, Stephan; Wang, Peng

I introduce a new approach to verifying fine-grained shared-memory concurrent programs modularly, not based on program logics. Rather, we define an instrumented operational semantics that includes fictitious code to watch all memory accesses and potentially signal a failure, embodying some formal protocol for object sharing. Several variants of the framework have been implemented in Coq at different levels of completeness, and one of our focuses is supporting mostly automated proofs for client code of intricate data structures.

3.3 A Calculus for Relaxed Memory

Karl Crary (Carnegie Mellon University, US)

License © Creative Commons BY 3.0 Unported license
© Karl Crary

Joint work of Crary, Karl; Sullivan, Michael J.

Main reference K. Crary, M. J. Sullivan, “A Calculus for Relaxed Memory,” in Proc. of the 42nd Annual ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages (POPL’15), pp. 623–636, ACM, 2015; pre-print available from author’s webpage.

URL <http://dx.doi.org/10.1145/2676726.2676984>

URL <http://www.cs.cmu.edu/~crary/papers/2015/rmc.pdf>

We propose a new approach to programming multi-core, relaxed-memory architectures in imperative, portable programming languages. Our memory model is based on explicit, programmer-specified requirements for order of execution and the visibility of writes. The

compiler then realizes those requirements in the most efficient manner it can. This is in contrast to existing memory models, which – if they allow programmer control over synchronization at all – are based on inferring the execution and visibility consequences of synchronization operations or annotations in the code.

We formalize our memory model in a core calculus called RMC. Outside of the programmer’s specified requirements, RMC is designed to be strictly more relaxed than existing architectures. It employs an aggressively nondeterministic semantics for expressions, in which actions can be executed in nearly any order, and a store semantics that generalizes Sarkar, *et al.*’s and Alglave, *et al.*’s models of the Power architecture. We establish several results for RMC, including sequential consistency for two programming disciplines, and an appropriate notion of type safety. All our results are formalized in Coq.

3.4 Modular Termination Verification for Non-blocking Concurrency

Pedro Da Rocha Pinto (Imperial College London, GB)

License  Creative Commons BY 3.0 Unported license
© Pedro Da Rocha Pinto

We present Total-TaDA, a program logic for verifying the total correctness of concurrent programs: that such programs both terminate and produce the correct result. The termination behaviour of a single thread can be conditional on the behaviour of its concurrent environment. With Total-TaDA, we are able to specify such constraints. This allows us to verify total correctness for non-blocking algorithms, such as a counter and a stack. Moreover, our approach is modular: we can verify the operations of a module independently, and build up modules on top of each other.

3.5 Speculation in Higher-Order Separation Logics (and why it’s tricky)

Thomas Dinsdale-Young (Aarhus University, DK)

License  Creative Commons BY 3.0 Unported license
© Thomas Dinsdale-Young
Joint work of Dinsdale-Young, Thomas; Svendsen, Kasper

When relating an implementation to an abstract specification, the abstract behaviours can depend on future concrete behaviours. We briefly motivate why this occurs, and consider how we might reason about a simple example in a modular way using a separation logic. Unfortunately, the naive approach is not sound, and we see why it leads to inconsistency.

3.6 Compositional C11 Program Transformation

Mike Dodds (University of York, GB)

License  Creative Commons BY 3.0 Unported license
© Mike Dodds
Joint work of Batty, Mark; Dodds, Mike; Gotsman, Alexey

One objective for language-level relaxed memory models is to support program transformations – i.e. compiler optimisations. However, it’s extremely subtle to calculate which

transformations are valid. This talk is about a theory for program transformations on the C11 relaxed model. Our theory is compositional: for each transformation, a limited number of executions represent all interactions with the context. To express these interactions, we use a partially-ordered record called a history (the set of histories could be seen as a kind of denotation). Our theory builds on ideas from C11 library abstraction: replacing a specification with an implementation is one instance of program transformation. This work is still in progress, but we already cover the core of the C11 model and many important transformations.

3.7 Static Verification of GPU Kernels

Alastair F. Donaldson (Imperial College London, GB)

License © Creative Commons BY 3.0 Unported license
© Alastair F. Donaldson

Joint work of Betts, Adam; Chong, Nathan; Donaldson, Alastair F.; Ketema, Jeroen; Qadeer, Shaz; Thomson, Paul; Wickerson, John

Main reference A. Betts, N. Chong, A. F. Donaldson, J. Ketema, S. Qadeer, P. Thomson, J. Wickerson, “The Design and Implementation of a Verification Technique for GPU Kernels,” *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 37(3):10:1–10:49, May 2015.

URL <http://dx.doi.org/10.1145/2743017>

During the presentation I gave a demonstration of GPUVerify, a static race-freedom verification tool for GPU kernels. GPUVerify enables scalable verification of massively parallel kernels through a combination of abstraction and sequentialization. Abstraction is applied to reduce the verification problem to the task of checking whether it is possible for two arbitrary threads to race. Sequentialization then exploits properties of the barrier-based GPU synchronization model so that verification for a pair of threads boils down to checking assertion-based correctness of a sequential program, whose size is linear in that of the source code for the original kernel (which itself is independent of the number of threads that execute the kernel).

This is joint work with the Multicore Programming Group at Imperial, and with Shaz Qadeer at Microsoft Research, and is described in a recent TOPLAS journal article.

References

- 1 Adam Betts, Nathan Chong, Alastair F. Donaldson, Jeroen Ketema, Shaz Qadeer, Paul Thomson, and John Wickerson. The design and implementation of a verification technique for GPU kernels. *ACM Trans. Program. Lang. Syst.*, 37(3):10:1–10:49, May 2015.

3.8 Making Sense of Rust (Work in Preservation)

Derek Dreyer (MPI-SWS – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
© Derek Dreyer

Joint work of Dreyer, Derek; Jung, Ralf; Turon, Aaron

Rust is a new language for “safe systems programming” developed at Mozilla, which uses an affine type system to guarantee type/memory safety and data race freedom. While the core type system is relatively simple and restrictive, essentially prohibiting aliased mutable state, many Rust libraries make significant internal use of “unsafe blocks” in order to escape this restriction. These uses of “unsafe” are supposedly encapsulated behind safe interfaces, but

they also fundamentally affect the meaning of types. For example, Rust libraries like `Rc` and `Cell` exhibit the phenomenon of “interior mutability”, whereby a supposedly “immutable” (roughly, “read-only”) operation on an object may in fact mutate its private state, so long as it does so in a way that does not violate the views of other aliases to the object. This has significant implications for concurrency, in particular leading to the need for a “Send” trait describing when a type is “thread-safe”. It’s also easy to get wrong, as evidenced by a recent soundness bug that was uncovered in the “scoped threads” API. In this work, which we have not yet even begun (!), we aim to develop a semantic model of Rust’s type system, based on Kripke logical relations and concurrent separation logic, which will enable us to make sense of what Rust types mean and to verify that the unsafe implementations of Rust libraries in fact preserve the end-to-end safety guarantees of the language.

3.9 An operational approach to relaxed memory models

Xinyu Feng (Univ. of Science & Technology of China – Suzhou, CN)

License  Creative Commons BY 3.0 Unported license

© Xinyu Feng

Joint work of Zhang, Yang; Feng, Xinyu

URL <http://staff.ustc.edu.cn/~xyfeng/research/publications/OHMM.html>

We present OHMM, an operational variation of the Happens-before Memory Model (HMM), the basis of Java memory model (JMM). OHMM is specified by giving an operational semantics to a language running on an abstract machine designed to simulate HMM. Thanks to its generative nature, the model naturally prevents out-of-thin-air reads. On the other hand, it uses a novel replay mechanism to allow instructions to be executed multiple times, which can be used to model many useful speculations and optimization. The model satisfies DRF- guarantee. It is weaker than JMM for lockless programs, thus can accommodate more optimization, such as the reordering of independent memory accesses that is not valid in JMM. Also many of the “ugly” examples in JMM are no longer ugly in our model. We hope OHMM can serve as the basis for new memory models for Java-like languages.

3.10 Formally Specifying POSIX File Systems

Philippa Gardner (Imperial College London, GB)

License  Creative Commons BY 3.0 Unported license

© Philippa Gardner

File system operations exhibit complex behaviour: they perform multiple actions affecting different parts of the state. This is further exacerbated when the operations are used concurrently. POSIX is a standard for operating systems, with a substantial part devoted to specifying file system operations. The specification is given in English, contains ambiguities and is generally under-specified with respect to concurrent behaviour. Therefore, it is not clear what clients may expect and what implementations must do. We extend modern concurrent program logics with a novel formalism for specifying multiple actions performed by an operation, which may be atomic, non-atomic or a combination of both, and give proof rules for client and implementation reasoning. With this formalism we give a formal specification to a common fragment of POSIX file system operations, and reason about clients such as lock files and an implementation of half-duplex pipes.

3.11 An Unsophisticated Higher-Order-ish Logic for Modular Specification and Verification of Total Correctness Properties of Fine-Grained Concurrent Imperative Programs

Bart Jacobs (KU Leuven, BE)

License © Creative Commons BY 3.0 Unported license
© Bart Jacobs

Many powerful higher-order logics have been proposed for the modular specification and verification of fine-grained concurrent imperative programs. In this talk, I present a logic that is fairly close to what my VeriFast modular verification tool for C and Java implements. To achieve higher-order-ishness (higher-order assertions, nested triples, assertions in the heap, etc.), a relatively simple approach is followed: assertion lambda applications and nested triples may occur only in positive positions. Negative facts can be passed around in the form of lemma lambdas, i.e. ghost command lambdas. We prove termination of such higher-order ghost code using call permissions, a technique we are presenting at ECOOP 2015 this summer.

3.12 Reasoning about possible values in concurrency

Cliff B. Jones (Newcastle University, GB)

License © Creative Commons BY 3.0 Unported license
© Cliff B. Jones

In joint research with Ian Hayes (Queensland) we are using a notation to express the ‘possible values’ of variables. So, for example, in a post condition of one process (we can not only talk about the initial and final values of a variable which might be changed by another process) – we can also specify in terms of the set of values that the environment might assign to a shared variable. Combined with rely/guarantee reasoning, this appears to offer clear and tractable specifications and reasoned designs. The possible values notation was shown on the example of Simpson’s four-slot implementation of Asynchronous Communication Mechanisms (ACMs).

3.13 Iris: Monoids and Invariants as an Orthogonal Basis for Concurrent Reasoning

Ralf Jung (MPI-SWS – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
© Ralf Jung

Joint work of Jung, Ralf; Swasey, David; Sieczkowski, Filip; Svendsen, Kasper; Turon, Aaron; Birkedal, Lars; Dreyer, Derek

Main reference R. Jung, D. Swasey, F. Sieczkowski, K. Svendsen, A. Turon, L. Birkedal, D. Dreyer, “Iris: Monoids and Invariants as an Orthogonal Basis for Concurrent Reasoning,” in Proc. of the 42nd Annual ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages (POPL’15), pp. 637–650, ACM, 2015.

URL <http://dx.doi.org/10.1145/2676726.2676980>

We present Iris, a concurrent separation logic with a simple premise: monoids and invariants are all you need. Partial commutative monoids enable us to express – and invariants enable

us to enforce – user-defined protocols on shared state, which are at the conceptual core of most recent program logics for concurrency. Furthermore, through a novel extension of the concept of a view shift, Iris supports the encoding of logically atomic specifications, i.e., Hoare-style specs that permit the client of an operation to treat the operation essentially as if it were atomic, even if it is not.

3.14 The Push/Pull Model of Transactions

Eric Koskinen (IBM TJ Watson Research Center – Yorktown Heights, US)

License © Creative Commons BY 3.0 Unported license
© Eric Koskinen

Joint work of Koskinen, Eric; Parkinson, Matthew

Main reference E. Koskinen, M. J. Parkinson, “The Push/Pull model of transactions,” in Proc. of the 36th ACM SIGPLAN Conf. on Programming Language Design and Implementation (PLDI’15), pp. 186–195, ACM, 2015; pre-print available from author’s webpage.

URL <http://dx.doi.org/10.1145/2737924.2737995>

URL <http://researcher.watson.ibm.com/researcher/files/us-ejk/pushpull.pdf>

We present a general theory of serializability, unifying a wide range of transactional algorithms, including some that are yet to come. To this end, we provide a compact semantics in which concurrent transactions PUSH their effects into the shared view (or UNPUSH to recall effects) and PULL the effects of potentially uncommitted concurrent transactions into their local view (or UNPULL to detangle). Each operation comes with simple criteria given in terms of commutativity (Lipton’s left-movers and right-movers).

The benefit of this model is that most of the elaborate reasoning (coinduction, simulation, subtle invariants, etc.) necessary for proving the serializability of a transactional algorithm is already proved within the semantic model. Thus, proving serializability (or opacity) amounts simply to mapping the algorithm on to our rules, and showing that it satisfies the rules’ criteria.

3.15 Curry-Howard for GUIs via Linear Temporal Classical Linear Logic

Neel Krishnaswami (University of Birmingham, GB)

License © Creative Commons BY 3.0 Unported license
© Neel Krishnaswami

Modern graphical user interface are structured with an event-driven architecture: programmers write programs as a collection of small imperative callbacks, which are invoked by an event loop as program events occur. That is, they must write higher-order imperative programs in continuation-passing style, which is notoriously challenging.

Using ideas from realizability theory, it is possible to build a model of classical linear logic on top of an event-based architecture. Since classical linear logic has a proof theory in terms of process calculi, we gain a neat explanation of why programmers talk about GUI programs in terms of concurrency, even though they implement them in terms of state and control. Furthermore, we now also have a type structure upon which we can build powerful abstractions – historically the bane of UI toolkits.

3.16 Owicki-Gries Reasoning for Weak Memory Models

Ori Lahav (MPI-SWS – Kaiserslautern, DE)

License © Creative Commons BY 3.0 Unported license
© Ori Lahav

Joint work of Lahav, Ori; Vafeiadis, Viktor

Main reference O. Lahav, V. Vafeiadis, “Owicki-Gries Reasoning for Weak Memory Models,” in Proc. of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP’15) – Part II, LNCS, Vol. 9135, pp. 311–323, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-662-47666-6_25

We show that even in the absence of auxiliary variables, the well-known Owicki-Gries method for verifying concurrent programs is unsound for weak memory models. By strengthening its non-interference check, however, we obtain OGRA, a program logic that is sound for reasoning about programs in the release-acquire fragment of the C11 memory model. We demonstrate the usefulness of this logic by applying it to several challenging examples, ranging from small litmus tests to an implementation of the RCU synchronization primitives.

3.17 A Program Logic for Contextual Refinement of Concurrent Objects under Fair Scheduling

Hongjin Liang (Univ. of Science & Technology of China – Suzhou, CN)

License © Creative Commons BY 3.0 Unported license
© Hongjin Liang

Joint work of Liang, Hongjin; Feng, Xinyu

Existing program logics on concurrent object verification either ignore progress properties, or aim for non-blocking progress (e.g., lock-freedom and wait-freedom), which cannot be applied to blocking algorithms that progress only under fair scheduling.

We present a new program logic for compositional verification of contextual refinement of concurrent objects under fair scheduling. As a key application, we show that starvation-freedom and linearizability of concurrent objects with blocking algorithms can be reformulated as contextual refinement, which can be verified using our program logic. With the logic, we have successfully verified starvation-freedom of simple algorithms using ticket locks, the two-lock queue algorithm and the lock-coupling list algorithm.

3.18 Formal Verification and Linux-Kernel Concurrency

Paul McKenney (IBM – Beaverton, US)

License © Creative Commons BY 3.0 Unported license
© Paul McKenney

Main reference P. E. McKenney, “Formal Verification and Linux-Kernel Concurrency,” presentation to Dagstuhl workshop 15191.

URL <http://materials.dagstuhl.de/files/15/15191/15191.PaulMcKenney.Slides.pdf>

This presentation reviews Linux-kernel validation, including its occasional use of formal verification, and presents conditions that a formal-verification tool would need to meet in order to be useful as part of the Linux kernel’s regression testing.

3.19 Linearizability: Who Really Needs It?

Paul McKenney (IBM – Beaverton, US)

License  Creative Commons BY 3.0 Unported license
© Paul McKenney

Main reference P. E. McKenney, “Linearizability: Who Really Needs It?,” Presentation to Dagstuhl 15191.
URL <http://materials.dagstuhl.de/files/15/15191/15191.PaulMcKenney1.Preprint.pdf>

Critique of the overuse of linearizability.

3.20 Some Examples of Kernel-Hacker Informal Correctness Reasoning

Paul McKenney (IBM – Beaverton, US)

License  Creative Commons BY 3.0 Unported license
© Paul McKenney

Main reference P. E. McKenney, “Some Examples of Kernel-Hacker Informal Correctness Reasoning,” Technical Report paulmck.2015.06.17a.
URL <http://www2.rdrop.com/users/paulmck/techreports/IntroRCU.2015.06.17a.pdf>

The examples include: (1) split counters, (2) RCU infrastructure, (3) RCU Small Bag use case, RCU Large Bag use case.

Also illustrates kernel-hacker reasoning surrounding RCU, along with one method of restoring consistency when using RCU. (Yes, there are other methods.)

3.21 Designing a Lock-Free Range Management Algorithm

Maged M. Michael (IBM TJ Watson Research Center – Yorktown Heights, US)

License  Creative Commons BY 3.0 Unported license
© Maged M. Michael

The talk describes the design process of a lock-free algorithm for allocation and deallocation of arbitrary large ranges. The algorithm is targeted to serve as a backend for known bounded-block-size lock-free memory allocators. It can serve as a user-level alternative to the `mmap` and `munmap` system calls in cases where the latter are unavailable or unsuitable. The algorithm aims to guarantee full coalescing. It uses only single word primitives: read, write, compare-and- swap; and it does not require any operating system calls. The algorithm supports continuous space availability, i.e., the space unavailable for allocation is bounded by the sum of allocated space and pending allocation requests.

3.22 Viper – A Verification Infrastructure for Permission based Reasoning

Peter Mueller (ETH Zürich, CH)

License © Creative Commons BY 3.0 Unported license
© Peter Mueller

Joint work of Juhasz, Uri; Kassios, Ioannis T.; Müller, Peter; Novacek, Milos; Schwerhoff, Malte; Summers, Alexander J.

Main reference U. Juhasz, I. T. Kassios, P. Müller, M. Novacek, M. Schwerhoff, A. J. Summers, “Viper: A Verification Infrastructure for Permission-Based Reasoning,” Unpublished manuscript.

URL <http://pm.inf.ethz.ch/publications/getpdf.php?bibname=Own&id=JKMNSS14.pdf>

The automation of verification techniques based on first-order logic specifications has benefited greatly from verification infrastructures such as Boogie and Why. These offer an intermediate language that can express diverse language features and verification techniques, as well as back-end tools such as verification condition generators.

However, these infrastructures are not well suited for verification techniques based on separation logic and other permission logics, because they do not provide direct support for permissions and because existing tools for these logics often prefer symbolic execution over verification condition generation. Consequently, tool support for these logics is typically developed independently for each technique, dramatically increasing the burden of developing automatic tools for permission-based verification.

In this talk, we present a verification infrastructure whose intermediate language supports an expressive permission model natively. We provide tool support, including two back-end verifiers, one based on symbolic execution, and one on verification condition generation; this facilitates experimenting with the two prevailing techniques in automated verification. Various existing verification techniques can be implemented via this infrastructure, alleviating much of the burden of building permission-based verifiers, and allowing the developers of higher-level techniques to focus their efforts at the appropriate level of abstraction.

3.23 Structures with Intrinsic Sharing, Subjectively

Aleksandar Nanevski (IMDEA Software – Madrid, ES)

License © Creative Commons BY 3.0 Unported license
© Aleksandar Nanevski

Joint work of Nanevski, Aleksandar; Ilya Sergey; Anindya Banerjee

The talk presents a new design pattern for proving correctness of data structures with deep sharing, such as graphs. The idea is to use subjective kind of auxiliary state, based on PCMs, which allows for threads to record their own changes to the datastructure, as well as the modifications performed by the interfering threads. The talk also discusses a rule for hiding, which introduces new auxiliary state within a delimited scope

3.24 An operational semantics for C/C++11 concurrency

Kyndylan Nienhuis (University of Cambridge, GB)

License  Creative Commons BY 3.0 Unported license
© Kyndylan Nienhuis

Joint work of Nienhuis, Kyndylan; Memarian, Kayvan; Pichon-Pharabod, Jean; Batty, Mark; Sewell, Peter

The axiomatic style of the C11 concurrency model makes it difficult to explore the possible execution of programs without exhaustive enumeration of all their candidate executions. Furthermore, since the rest of C is defined in an operational style, it is difficult to extend the concurrency model to a semantics for the whole language.

We present ongoing research on an operational concurrency model for C11 that is equivalent to the axiomatic model, executable, and integratable with an operational semantics for sequential C. This work also reveals omissions in the definition of C: notions such as lifetime and undefined behaviour are defined for sequential C only, and we discovered that their definitions do not generalise to concurrent C.

3.25 Investigating Weak Memory Performance

Scott Owens (University of Kent, GB)

License  Creative Commons BY 3.0 Unported license
© Scott Owens

Joint work of Ritson, Carl; Owens, Scott

This talk will describe some preliminary and ongoing work into the real-world performance implications of fence placement strategies on ARM and POWER architectures.

3.26 Polarized Substructural Session Types

Frank Pfenning (Carnegie Mellon University, US)

License  Creative Commons BY 3.0 Unported license
© Frank Pfenning

We provide an overview of session-typed message-passing concurrent programming, which arises from a Curry-Howard interpretation of (intuitionistic) linear logic. Most recent work considers multiple structural properties (linear, affine, and unrestricted) connected by modal operators. The same modal operators (often called “up” and “down”) can also be used to mediate between positive and negative linear proposition, one corresponding to output and one to input.

References

- 1 Frank Pfenning and Dennis Griffith. Polarized substructural session types. In A. Pitts, editor, *Proceedings of the 18th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS 2015)*, pages 3–22, London, England, April 2015. Springer LNCS 9034. Invited talk.

3.27 An attempt at fixing C11 concurrency

Jean Pichon-Pharabod (University of Cambridge, GB)

License © Creative Commons BY 3.0 Unported license
© Jean Pichon-Pharabod

The memory model of the C programming language, which defines what values a read in a concurrent program can read, allows reads to read value that are not constructed by the program, but appear “out of thin air”. We argue that this problem is due to the memory model considering the wrong objects, namely configurations in the naive event structure of the program. We propose an alternative memory model for locks and non-atomic and relaxed accesses based on considering the whole event structure.

3.28 Automated and Modular Refinement Reasoning for Concurrent Programs

Shaz Qadeer (Microsoft Corporation – Redmond, US)

License © Creative Commons BY 3.0 Unported license
© Shaz Qadeer

Joint work of Hawblitzel, Chris; Petrank, Erez; Qadeer, Shaz; Tasiran, Serdar

Main reference C. Hawblitzel, E. Petrank, S. Qadeer, S. Tasiran, “Automated and Modular Refinement Reasoning for Concurrent Programs,” in Proc. of the 27th International Conference on Computer Aided Verification (CAV’15), LNCS, Vol. 9207, pp. 449–465, Springer, 2015; preliminary technical report available, MSR-TR-2015-8, Microsoft Research, 2015.

URL http://dx.doi.org/10.1007/978-3-319-21668-3_26

URL <http://research.microsoft.com/apps/pubs/?id=238907>

URL <http://research.microsoft.com/apps/pubs/?id=258112>

We present CIVL, a language and verifier for concurrent programs based on automated and modular refinement reasoning. CIVL supports reasoning about a concurrent program at many levels of abstraction. Atomic actions in a high-level description are refined to fine-grain and optimized lower-level implementations. Modular specifications and proof annotations, such as location invariants and procedure pre- and post-conditions, are specified separately, independently at each level in terms of the variables visible at that level. We have implemented CIVL as an extension to the Boogie language and verifier. We have used CIVL to refine a realistic concurrent garbage collection algorithm from a simple high-level specification down to a highly-concurrent implementation described in terms of individual memory accesses.

3.29 CoLoSL: Concurrent Local Subjective Logic

Azalea Raad (Imperial College London, GB)

License © Creative Commons BY 3.0 Unported license
© Azalea Raad

Joint work of Raad, Azalea; Villard, Jules; Gardner, Philippa

Main reference A. Raad, J. Villard, P. Gardner, “CoLoSL: Concurrent Local Subjective Logic,” in Proc. of the 24th European Symposium on Programming on Programming Languages and Systems (ESOP’15), LNCS, Vol. 9032, pp. 710–735, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-662-46669-8_29

A key difficulty in verifying shared-memory concurrent programs is reasoning compositionally about each thread in isolation. Existing verification techniques for fine-grained concurrency

typically require reasoning about either the entire shared state or disjoint parts of the shared state, impeding compositionality. In this work we introduce the program logic CoLoSL, where each thread is verified with respect to its subjective view of the global shared state. This subjective view describes only that part of the state accessed by the thread. Subjective views may arbitrarily overlap with each other, and expand and contract depending on the resource required by the thread. This flexibility gives rise to small specifications and, hence, more compositional reasoning for concurrent programs. We demonstrate our reasoning on a range of examples, including a concurrent computation of a spanning tree of a graph.

3.30 Concurrency-Aware Linearizability

Noam Rinetzky (Tel Aviv University, IL)

License © Creative Commons BY 3.0 Unported license
© Noam Rinetzky

Joint work of Hemed, Nir; Rinetzky, Noam

Main reference N. Hemed, N. Rinetzky, “Brief Announcement: Concurrency-Aware Linearizability,” in Proc. of the 2014 ACM Symp. on Principles of Distributed Computing (PODC’14), pp. 209–211, ACM, 2014; pre-print available from author’s webpage.

URL <http://dx.doi.org/10.1145/2611462.2611513>

URL <http://www.cs.tau.ac.il/~maon/pubs/podc14.pdf>

Linearizability allows to describe the behaviour of concurrent objects using sequential specifications. Unfortunately, as we show in this paper, sequential specifications cannot be used for concurrent objects whose observable behaviour in the presence of concurrent operations should be different than their behaviour in the sequential setting. As a result, such concurrency-aware objects do not have formal specifications, which, in turn, precludes formal verification.

In this paper we present Concurrency Aware Linearizability (CAL), a new correctness condition which allows to formally specify the behaviour of a certain class of concurrency-aware objects. Technically, CAL is formalized as a strict extension of linearizability, where concurrency-aware specifications are used instead of sequential ones. We believe that CAL can be used as a basis for modular formal verification techniques for concurrency-aware objects.

3.31 Anatomy of mechanized reasoning about fine-grained concurrency

Ilya Sergey (IMDEA Software – Madrid, ES)

License © Creative Commons BY 3.0 Unported license
© Ilya Sergey

URL <http://ilyasergey.net/slides/2015-Sergey-al-Dagstuhl.pdf>

In this talk, I will give a quick hands-on demo, explaining the structure of the proofs when verifying fine-grained concurrent programs in the recently proposed Coq-based framework of Fine-grained Concurrent Separation Logic.

I will outline key stages of formalization of characteristic concurrent protocols, explaining the encoding of atomic actions and stable specifications. I will also outline typical proof patterns, appearing during the reasoning about composition of concurrent specifications.

3.32 Using Iris as a meta-language for logical relations

Kasper Svendsen (Aarhus University, DK)

License © Creative Commons BY 3.0 Unported license
© Kasper Svendsen

Joint work of Svendsen, Kasper; Birkedal, Lars; Askarov, Aslan; Krog-Jespersen, Morten

In this talk, I argue that Iris is well-suited as a meta-language for defining binary Step-indexed Kripke Logical Relations. Step-indexed Kripke Logical Relations provide a very powerful proof technique for reasoning about realistic languages. However, they can be difficult to define and work with directly, requiring explicit reasoning about steps and the existence of recursively-defined Kripke worlds. Using Iris as a meta-language, we can hide the steps and avoid the construction of recursively-defined worlds, by piggy-backing on Iris' impredicative invariants and monoids.

3.33 Verifying Read-Copy-Update in a Logic for Weak Memory

Joseph Tassarotti (Carnegie Mellon University, US)

License © Creative Commons BY 3.0 Unported license
© Joseph Tassarotti

Joint work of Tassarotti, Joseph; Dreyer, Derek; Vafeiadis, Viktor

Main reference J. Tassarotti, D. Dreyer, V. Vafeiadis, “Verifying read-copy-update in a logic for weak memory,” in Proc. of the 36th ACM SIGPLAN Conf. on Programming Language Design and Implementation (PLDI'15), pp. 110–120, ACM, 2015.

URL <http://dx.doi.org/10.1145/2737924.2737992>

Read-Copy-Update (RCU) is a technique for letting multiple readers safely access a data structure while a writer concurrently modifies it. It is used heavily in the Linux kernel in situations where fast reads are important and writes are infrequent. Optimized implementations rely only on the weaker memory orderings provided by modern hardware, avoiding the need for expensive synchronization instructions (such as memory barriers) as much as possible.

Using GPS, a recently developed program logic for the C/C++11 memory model, we verify an implementation of RCU for a singly-linked list assuming “release-acquire” semantics. Although release-acquire synchronization is stronger than what is required by real RCU implementations, it is nonetheless significantly weaker than the assumption of sequential consistency made in prior work on RCU verification. Ours is the first formal proof of correctness for an implementation of RCU under a weak memory model

3.34 Software verification under weak memory consistency

Viktor Vafeiadis (MPI-SWS – Kaiserslautern, DE)

License © Creative Commons BY 3.0 Unported license
© Viktor Vafeiadis

Weak memory consistency makes reasoning about concurrent programs rather challenging as it invalidates many of the traditional reasoning techniques that are sound under sequential consistency. The talk demonstrates some of the challenges involved and possible solutions.

3.35 Open Problems and State-of-Art of Session Types

Nobuko Yoshida (Imperial College London, GB)

License © Creative Commons BY 3.0 Unported license
© Nobuko Yoshida

Main reference R. Demangeon, K. Honda, R. Hu, R. Neykova, N. Yoshida, “Practical interruptible conversations: distributed dynamic verification with multiparty session types and Python,” *Formal Methods in System Design*, 46(3):197–225, 2015.

URL <http://dx.doi.org/10.1007/s10703-014-0218-8>

We give a summary of our recent research developments on multiparty session types for verifying distributed and concurrent programs, and our collaborations with industry partners and a major, long-term, NSF-funded project (Ocean Observatories Initiatives) to provide an ultra large-scale cyberinfrastructure (OOI CI) for 25-30 years of sustained ocean measurements to study climate variability, ocean circulation and ecosystem dynamics. We shall first talk how Robin Milner, Kohei Honda and Yoshida started collaborations with industry to develop a web service protocol description language called Scribble and discovered the theory of multiparty session types through the collaborations. We then talk about the recent developments in Scribble and the runtime session monitoring framework currently used in the OOI CI.

Participants

- Lennart Beringer
Princeton University, US
- Lars Birkedal
Aarhus University, DK
- Andrea Cerone
IMDEA Software – Madrid, ES
- Adam Chlipala
MIT – Cambridge, US
- Karl Crary
Carnegie Mellon University, US
- Pedro Da Rocha Pinto
Imperial College London, GB
- Thomas Dinsdale-Young
Aarhus University, DK
- Mike Dodds
University of York, GB
- Alastair F. Donaldson
Imperial College London, GB
- Cezara Dragoi
IST Austria –
Klosterneuburg, AT
- Derek Dreyer
MPI-SWS – Saarbrücken, DE
- Xinyu Feng
Univ. of Science & Technology of
China – Suzhou, CN
- Philippa Gardner
Imperial College London, GB
- Alexey Gotsman
IMDEA Software – Madrid, ES
- Aquinas Hobor
National Univ. of Singapore, SG
- Jan Hoffmann
Yale University, US
- Chung-Kil Hur
Seoul National University, KR
- Bart Jacobs
KU Leuven, BE
- Cliff B. Jones
Newcastle University, GB
- Ralf Jung
MPI-SWS – Saarbrücken, DE
- Eric Koskinen
IBM TJ Watson Research Center
– Yorktown Heights, US
- Neel Krishnaswami
University of Birmingham, GB
- Ori Lahav
MPI-SWS – Kaiserslautern, DE
- Hongjin Liang
Univ. of Science & Technology of
China – Suzhou, CN
- Paul McKenney
IBM – Beaverton, US
- Maged M. Michael
IBM TJ Watson Res. Center –
Yorktown Heights, US
- Peter Müller
ETH Zürich, CH
- Aleksandar Nanevski
IMDEA Software – Madrid, ES
- Kyndylan Nienhuis
University of Cambridge, GB
- Scott Owens
University of Kent, GB
- Frank Pfenning
Carnegie Mellon University, US
- Jean Pichon-Pharabod
University of Cambridge, GB
- Francois Pottier
INRIA – Le Chesnay, FR
- Shaz Qadeer
Microsoft Corporation –
Redmond, US
- Azalea Raad
Imperial College London, GB
- John Reppy
University of Chicago, US
- Noam Rinetzky
Tel Aviv University, IL
- Claudio Russo
Microsoft Research UK –
Cambridge, GB
- Ilya Sergey
IMDEA Software – Madrid, ES
- Peter Sewell
University of Cambridge, GB
- Zhong Shao
Yale University, US
- Kasper Svendsen
Aarhus University, DK
- Joseph Tassarotti
Carnegie Mellon University, US
- Viktor Vafeiadis
MPI-SWS – Kaiserslautern, DE
- Martin T. Vechev
ETH Zürich, CH
- Nobuko Yoshida
Imperial College London, GB



The Message in the Shadow: Noise or Knowledge?

Edited by

Roberto Casati¹, Patrick Cavanagh², and Paulo E. Santos³

1 CNRS – Paris, FR, roberto_casati@ens.fr

2 Paris Descartes University, FR, patrick.cavanagh@parisdescartes.fr

3 Centro Universitario da FEI – São Paulo, BR, psantos@fei.edu.br

Abstract

Computer vision, besides being a key area in Computer Science, is present in various industrial applications, such as traffic sign recognition (including car license plates), face and gesture recognition, content-based image retrieval, remote sensing, cartography, radar sensing, and robot mapping. However, most computer vision systems disregard the cognitive aspects of human perception, thus limiting their applicability in natural environments, whereby small changes in the light conditions cause negative effects on the system's accuracy. This seminar brought together contributions from Computer Vision, Cognitive Psychology, Philosophy and History of Art in order to discuss the information content in cast shadows which, although currently recognised by psychologists as providing important cues about depth perception, is considered as noise in the computer vision literature.

Seminar May 3–8, 2015 – <http://www.dagstuhl.de/15192>

1998 ACM Subject Classification I.2.10 Visual and Scene Understanding; I.4.8 Scene Analysis; J.5 Arts and Humanities.

Keywords and phrases Cast Shadows, Perception, Computer Vision, Space Cognition

Digital Object Identifier 10.4230/DagRep.5.5.24

Edited in cooperation with Valquiria Fenelon

1 Executive Summary

Roberto Casati

Patrick Cavanagh

Paulo E. Santos

License  Creative Commons BY 3.0 Unported license
© Roberto Casati, Patrick Cavanagh, and Paulo E. Santos

The seminar “The Message in the Shadow: Noise or Knowledge?” brought together researchers from the various disciplines involved in investigating the problem of understanding the perception of shadows (both in biological and in artificial systems) as well as art historians and artists involved in the study or in the manipulation of shadows in art pieces. The nationalities of the seminar participants were as varied as the disciplines involved its central theme; from the 20 attendees there were 4 that came from Brazil, 4 from Germany, 1 from the Emirates, 2 from France, 2 from the UK, 1 from Canada, 3 from the US, 1 from the Netherlands and 2 from Japan.

The small size of the seminar helped to create a friendly atmosphere, in which every participant had time and space to engage in discussions with every other, and every one had an equal amount of time to present his/her ideas, independently of the career stage the participant was in.



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

The Message in the Shadow: Noise or Knowledge?, *Dagstuhl Reports*, Vol. 5, Issue 5, pp. 24–42

Editors: Roberto Casati, Patrick Cavanagh, and Paulo E. Santos



DAGSTUHL
REPORTS

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

The dynamics of the seminar was as follows: every participant that had an interest in presenting a talk was allocated a 20 minute slot, followed by a 10 min discussion period, during the mornings (from 9 to 11am). The talks were distributed into 4 tracks, one for each day of the week: Psychology (Monday), Artificial Intelligence and Computer Vision (Tuesday), Art and Rendering (Wednesday), Architecture and Spatial Reasoning (Thursday). The titles of the talks given, per track, are cited as follows (the related abstracts are listed in the next section):

Psychology

- Patrick Cavanagh, What does vision know about shadows?
- John Kennedy, Shape-from-shadow polarity
- John O’Dea, Do shadows make surfaces look dark?
- Marteen Wijntjes, Perception of shadows in paintings

Artificial Intelligence and Computer Vision

- Hannah M. Dee, Why does computer vision find shadows so problematic?
- Paulo E. Santos, Shadows in AI and Robotics
- Frederick Fol Leymarie, On medialness-based shape representation: recent developments and food for thought
- Ann Marie Raynal, Leveraging the Information in the Shadows of Synthetic Aperture Radar

Art and Rendering

- Koichi Toyama, The systematic introduction of Chiaroscuro in 15th century Florence and the symbolic shadow in Sieneese Painting
- William Sharpe, Shadow Messages in the arts
- Marcos Danhoni, Shadows on the moon and the sun by Cigoliand Galileo: The Copernican planetarium inside the Paolina ´s Chapel of Santa Maria Maggiore
- Roberto Casati, X-From-Shadow: There is still room at the bottom
- Koichi Toyama, Un-naturalistic painting and the lack of shadow: History of shadow in 18th- 19th century Japanese paintings and woodblock prints

Architecture and Spatial Reasoning

- Barbara Tversky, Can uses of shadows in language and art inform perception of shadows?
- Juliano Beraldo, Daylight metrics for building design
- Christian Freksa, Shadow and friends illuminate space
- Mehul Bhatt, Carl Schultz and Jakob Suchan, Grasping Objectified Shadows

Working Groups

At the end of the morning sessions, discussions were conducted in which the ideas presented during the talks served as inspiration for the conception of research statements. Some of these statements were selected to be discussed during the break out session that occurred during the Monday and Tuesday afternoons. The main questions discussed are presented below:

- Information about the light-source contained in shadows: there is a number of features from the light source that is present in the shadow of an object (for instance: the number of sources, the localisation, the shape) but much of this information is not used by the perceptual system. The question of the evolutionary advantages of this selective use of the information content of shadows was discussed and also the possibilities for a computer system to explore it fully;
- Mooney Faces and Shadows: To test people's vision, Craig Mooney devised two-tone pictures of faces. In Mooney faces, some parts are strongly illuminated, others are in deep shadow. His pictures were static. Motion helps vision find the faces. Mooney faces in negatives are hard to make out. Proper facial expression is lost. In outline, they are equally uninterpretable. Adding a dark line to the border of a positive Mooney face can drop recognition to the level of a negative. Motion helps, but still leaves the face looking cartoonish and flat. Often the line is taken as part of a profile. A light line border of a negative also leaves it cartoonish.
- Cross-disciplinary terminology for shadows: there is currently a non-consensual use of terms to refer to shadow issues (for instance, a caster is sometimes referred to as 'obtruder' or 'occluder'). This group proposed a tentative terminology that was later discussed with the other participants.
- Throwing away information. Shadows are used by the visual system to retrieve various spatial features of the scene, then discarded. The group discussed cognitive/computational mechanisms that may throw away shadows.
- Mereotopological formalisation of Eclipses. The group created a formalized version of the terminology used in describing the different phases of an Eclipse of the Sun. An amendment of the existing taxonomy was proposed.

At the end of the Monday session, artist Francesca Bizzarri showed some aspects of the art of shadow performance.

On Thursday afternoon the participants were directed to discuss possible collaborations, project proposals, and to devise conclusions (even if partial) to the various questions discussed during the previous days. Some of the results obtained in this session are listed below:

- Collaboration between S. Paulo and Bremen
- Online, real-time, Mooney face generator – A computer generated video by Dee, Kennedy and Casati, on the impairment of depth perception through the display of lines on moving Mooney faces, has been created and is visible at:
<https://www.youtube.com/watch?v=IuDNUz9RSuw>
- Collaboration between Tokyo and New York (on art history)
- The foundation of a work group on terminology
- The projected publication of the mereotopological formalisation of Eclipses (Paris-Bremen-S. Paulo)
- Video displaying the phenomenon of the polarization of shadow (Casati and Cavanagh)

Finally, we discussed the future submission of a proposal for a special issue of the Journal Spatial Cognition and Computation (<http://www.tandfonline.com/toc/hsc20/current>) with the themes of the seminar and the organisation of a follow-up event in 2017 related to these ideas. Our proposal for a special issue of the Journal Spatial Cognition and Computation was accepted by the journal editors in June, 2015 (the *call for papers* will be advertised in the second half of this year).

2 Table of Contents

Executive Summary

<i>Roberto Casati, Patrick Cavanagh, and Paulo E. Santos</i>	24
--	----

Overview of Talks

Daylight metrics for building design <i>Juliano Beraldo</i>	29
Declarative Reasoning about Spatio-Temporal Things <i>Mehul Bhatt</i>	30
Ideas on shadow theatre <i>Francesca Bizzarri</i>	32
X-from-Shadows: There is still room at the bottom <i>Roberto Casati</i>	32
What does vision know about shadows? <i>Patrick Cavanagh</i>	33
Shadows on the Moon and the Sun by Cigoli and Galilei: The Copernican planetarium inside Paolina's Chapel of Santa Maria Maggiore. <i>Marcos Cesar Danhoni Neves</i>	33
Why does computer vision find shadows so problematic? <i>Hannah M. Dee</i>	34
Shadows, Shape, Medialness <i>Frederic Fol Leymarie</i>	34
Shadow and friends illuminate space <i>Christian Freksa</i>	35
Shape-from-shadow polarity applies to 11 borders – 8 real, 3 pictorial <i>John Kennedy</i>	35
Do Shadows Make Surfaces look Dark? <i>John O'Dea</i>	36
Leveraging the Information in the Shadows of Synthetic Aperture Radar <i>Ann Marie Raynal</i>	36
Shadows in Robotics and Artificial Intelligence <i>Paulo E. Santos</i>	38
Grasping Objectified Shadows <i>Carl Schultz</i>	38
Shadow Messages in the Arts <i>William Sharpe</i>	38
Declarative Methods for Cognitive Vision <i>Jakob Suchan</i>	39
The systematic introduction of Chiaroscuro in 15th century Florence and the symbolic shadow in Sieneese Painting and Un-naturalistic painting <i>Koichi Toyama</i>	40

Un-naturalistic painting and the lack of shadow: History of shadow in 18th-19th century Japanese paintings and woodblock prints	
<i>Koichi Toyama</i>	40
Shadow Play	
<i>Barbara Tversky</i>	41
Perception of Shadow and Shading in Paintings and Photographs	
<i>Maarten Wijntjes</i>	41
Participants	42

3 Overview of Talks

3.1 Daylight metrics for building design

Juliano Beraldo (University São Paulo, BR)

License © Creative Commons BY 3.0 Unported license
© Juliano Beraldo

Main reference International Commission on Illumination (CIE) – <http://www.cie.co.at>, Sessions of the CIE, Expert Symposium

Main reference D. DiLaura, K. Houser, R. Mistrick, G. Steffy (eds.), “The Lighting Handbook,” 10th Edition, Illuminating Engineering Society of North America (IESNA).

Main reference T. Muneer, C. Gueymard, H. Kambezidis, “Solar Radiation and Daylight Models,” 2nd edition, Elsevier Butterworth-Heinemann, 2004.

Main reference J. Summerson, “The Classical Language of Architecture,” ISBN: 9780262690126, MIT Press, 1966.

This presentation covers the findings of the research that is still ongoing and shows the daylighting traditional project approach and new daylight performance metrics. Also discuss the interaction between architectural and daylighting project aspects and finish with a reflective discussion about the project approach in which the shadow is understood not only as a variation on intensity of light but also as a variation in the light quality.

The boom in the construction of energy efficient buildings promoted the use of solar architectural features to maximize the exploitation of daylight. Nowadays, recent policies regarding energy performance of buildings have led to vigorous standardisation activities. Rating systems, such as LEED, BREEAM, DGNB, are now a key driver of building design. However, daylighting is a building performance strategy, which is difficult to evaluate due to the fact that embraces several different dynamic interaction factors from building, climate and surroundings.

The current design and evaluation of daylighting approaches revealed to be unsatisfactory faced with the new designers evaluation requirements of a more complete approach. As a result, during the last decade, there has been several researches related to daylighting and new empirical models have been developed.

In the last years the use of dynamic daylight performance measures was promoted by CIE (Commission Internationale de l’Eclairage) as capable to lead to superior daylighting designs. Nevertheless, recent discussions in CIE sessions pointed to the need of verifying the many dimensions of daylighting performances has revealed that the current evaluation methods and metrics in many occasions proved to be deficient or inadequate to assess daylighting in a holistic way.

The changes in the methods do not change the traditional principles of good daylighting practice that are focus in three main points: light levels, indoors light distribution and avoid glare. However, an overall point of view is essential, according to latest scientific debates, to achieve the high-quality building environment. It is possible that a small change in the daylighting project approach, based on geometric proportions and integrated with an architectural perspective, could reach a more appropriate daylighting results.

3.2 Declarative Reasoning about Spatio-Temporal Things

Mehul Bhatt (*Universität Bremen, DE*)

License © Creative Commons BY 3.0 Unported license
© Mehul Bhatt

Joint work of Bhatt, Mehul; Schultz, Carl; Suchan, Jakob

Main reference M. Bhatt, “Between Sense and Sensibility: Declarative narrativisation of mental models as a basis and benchmark for visuo-spatial cognition and computation focussed collaborative cognitive systems,” arXiv:1307.3040v2 [cs.AI], 2104.

URL <http://arxiv.org/abs/1307.3040v2>

We pursue declarative reasoning about *visuo-spatial dynamics* from the viewpoint of the research areas of *artificial intelligence*, *commonsense reasoning*, and *spatial cognition and computation*.

Declarative spatial reasoning ([3]) is the ability to (declaratively) specify and solve real-world problems related to geometric and qualitative visuo-spatial representation and reasoning pertaining to spatial, temporal, and spatio-temporal things, be it abstract regions of space or time, material spatial objects, or spatial artefacts such as shadows or affordances cast by material objects ([2, 4, 10]). The problems that we address in this context encompass both specialist and everyday instances identifiable in a range of *cognitive technologies* and *spatial assistance systems* where spatio-linguistic conceptualisation & background knowledge focussed visuo-spatial cognition and computation are central ([5]).

As a systematic development of the declarative spatial reasoning method, we have initiated formalisations of *space* and *spatial reasoning* within constraint logic programming ([3, 8, 9]), and most recently, non-monotonic reasoning about spatio-temporal dynamics within answer set programming ([12]). We have developed **CLP(QS)**, a *declarative spatial reasoning system* capable of modelling and reasoning about qualitative spatial relations pertaining to multiple spatial domains, i.e., one or more aspects of space such as topology, and intrinsic and extrinsic orientation, size, distance etc. With CLP(QS), users and application developers may freely mix object domains (i.e., points, line-segments, and regions) with the available spatial domains. CLP(QS) also offers mixed geometric-qualitative spatial reasoning capabilities, and in its current form, basic quantification support offering the means to go back from qualitative relations to the domain of precise quantitative information.

The emphasis in CLP(QS) is on the seamless integration of declarative visuospatial (computational) problem-solving capabilities within large-scale hybrid AI systems, and cognitive (interaction) technologies. Currently, integration is achieved via the medium of *logic programming* – specifications in the form of (domain) *facts* and *rules* consisting of mix of, for instance, background semantic or conceptual knowledge, spatio-temporal knowledge, and knowledge about action and dynamics. The general concept of declarative spatial reasoning lends itself to re-interpretations and extensions with other perspectives such as diagrammatic representations. CLP(QS) marks a clear departure from other (relational-algebraically based) spatial reasoning methods / tools by its use of the constraint logic programming framework for formalising the semantics of mixed geometric and qualitative spatial representation and reasoning. The approach has demonstrated applicability in several domains, most recent examples being *architectural design cognition* ([7]), *cognitive vision* ([6, 11]), *geospatial information systems* [1].

The CLP(QS) system is also being designed and used as a pedagogical tool to be used as part of university based courses at the interface of Artificial Intelligence, Knowledge Representation and Reasoning, Cognitive Systems, and Spatial Informatics.

CLP(QS): A Declarative Spatial Reasoning System. <http://www.spatial-reasoning.com>.

References

- 1 M. Bhatt and J. O. Wallgruen. Geospatial narratives and their spatio-temporal dynamics: Commonsense reasoning for high-level analyses in geographic information systems. *ISPRS International Journal of Geo-Information., Special Issue on: Geospatial Monitoring and Modelling of Environmental Change*, 3(1):166–205, 2014. ISSN 2220-9964. 10.3390/ijgi3010166. URL <http://www.mdpi.com/2220-9964/3/1/166>.
- 2 M. Bhatt, F. Dylla, and J. Hois. Spatio-terminological inference for the design of ambient environments. In *Spatial Information Theory, 9th Int'l Conf., COSIT 2009, Aber Wrac'h, France, September 21-25, 2009, Proceedings*, volume 5756 of *LNCS*, pages 371–391. Springer, 2009. 10.1007/978-3-642-03832-7_23. URL http://dx.doi.org/10.1007/978-3-642-03832-7_23.
- 3 M. Bhatt, J. H. Lee, and C. Schultz. CLP(QS): A Declarative Spatial Reasoning Framework. In *Proceedings of the 10th Int'l Conf. on Spatial information theory, COSIT'11*, pages 210–230, Springer-Verlag, 2011. ISBN 978-3-642-23195-7.
- 4 M. Bhatt, C. P. L. Schultz, and M. Huang. The shape of empty space: Human-centred cognitive foundations in computing for spatial design. In *2012 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC 2012, Innsbruck, Austria, September 30 – October 4, 2012*, pages 33–40. IEEE, 2012. ISBN 978-1-4673-0852-6. 10.1109/VLHCC.2012.6344477. <http://dx.doi.org/10.1109/VLHCC.2012.6344477>.
- 5 M. Bhatt, C. Schultz, and C. Freksa. The ‘Space’ in Spatial Assistance Systems: Conception, Formalisation and Computation. In T. Tenbrink, J. Wiener, and C. Claramunt, editors, *Representing space in cognition: Interrelations of behavior, language, and formal models. Series: Explorations in Language and Space*, Explorations in Language and Space. 978-0-19-967991-1, Oxford University Press, 2013a. ISBN 9780199679911.
- 6 M. Bhatt, J. Suchan, and C. Schultz. Cognitive Interpretation of Everyday Activities – Toward Perceptual Narrative Based Visuo-Spatial Scene Interpretation. In M. Finlayson, B. Fisseni, B. Loewe, and J. C. Meister, editors, *Computational Models of Narrative (CMN) 2013., a satellite workshop of CogSci 2013: The 35th meeting of the Cognitive Science Society.*, OpenAccess Series in Informatics (OASiCs), Schloss Dagstuhl, Germany, 2013b. .
- 7 M. Bhatt, C. P. L. Schultz, and M. Thosar. Computing narratives of cognitive user experience for building design analysis: KR for industry scale computer-aided architecture design. In C. Baral, G. D. Giacomo, and T. Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proc. of the Fourteenth Int'l Conf., KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014. ISBN 978-1-57735-657-8.
- 8 C. Schultz and M. Bhatt. Towards a Declarative Spatial Reasoning System. In *ECAI 2012 – 20th European Conference on Artificial Intelligence., Montpellier, France, 2012*.
- 9 C. Schultz and M. Bhatt. Declarative spatial reasoning with boolean combinations of axis-aligned rectangular polytopes. In *21st Europ. Conf. on Artificial Intelligence (ECAI'14)*, 2014.
- 10 C. P. L. Schultz and M. Bhatt. Declarative computing with shapes and shadows. In *Proceedings of the Second Interdisciplinary Workshop The Shape of Things, Rio de Janeiro, Brazil, April 3-4, 2013*, volume 1007 of *CEUR Workshop Proceedings*, pages 143–150. CEUR-WS.org, 2013. <http://ceur-ws.org/Vol-1007/paper12.pdf>.
- 11 J. Suchan, M. Bhatt, and P. E. Santos. Perceptual narratives of space and motion for semantic interpretation of visual data. In *Computer Vision – ECCV 2014 Workshops – Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part II*, volume 8926 of *LNCS*, pages 339–354. Springer, 2014. http://dx.doi.org/10.1007/978-3-319-16181-5_24.
- 12 P. Walega, M. Bhatt, and C. Schultz. ASPMT(QS): Non-Monotonic Spatial Reasoning with Answer Set Programming Modulo Theories. In *LPNMR: Logic Programming and Nonmonotonic Reasoning – 13th International Conference, 2015*.

3.3 Ideas on shadow theatre

Francesca Bizzarri (Associazione Ca' Luogo D'Arte – Gattatico, IT)

License  Creative Commons BY 3.0 Unported license
© Francesca Bizzarri

In theatre: pretense is more real than reality and sometimes the shadow in theatre helps the pleasure of imagining; shadow is ambiguous, we complete the emotion. In theatre, as a condensation of time, space and emotion, the use of shadow engage audience to fill in shadow blanks and supports some technical possibilities, not always doable in real: Alice in Wonderland grows and shrinks easily. Usually an actor has to adjust action speed to shadow speed; style and angle of gesture are important; feel the light in order to know you are projecting a profile; know where your body is vis à vis the light; an actor has to imagine what the audience is seeing/imagining. In some shadow theatre we do not see features, so the body and gesture are important; but in regular theatre the lighting technician has to cancel distracting shadows; as actors we have to “grab the light”, we spend one day to pointing the light on all the scenes.

In Noh Theatre the changing shadow on a mask gives different emotions and expressions. Moving shadows convey action, gender, the exact person if you know him/her. When in theatre halogen lamp came out every shadow became sharp, clean, readable. Transformations, creating illusions depends on shadows because you only get the outline, you lose body and individuality; creation of airplanes or flowers; modern shadow theatre companies use whole troupe to make shapes, objects, setting, not just human figures and they rely on the astonishment of the audience, the surprise to seeing people represent complete different figures even inanimate ones. When we see shadows on screen do we see them as flat, as figures in a flat world, or as three dimensional? Do we fill in a third dimension, do we imagine colors, materials?

3.4 X-from-Shadows: There is still room at the bottom

Roberto Casati (ENS – Paris, FR)

License  Creative Commons BY 3.0 Unported license
© Roberto Casati

I shall present a number of ecological observations and pictorial examples to indicate that there is still room at the bottom for research in shadow perception. Shadows are minor entities, both ontologically and perceptually, and the examples presented here are thus marginal cases of a minor entity. Not all the information present in shadows appears to be retrieved by the visual system (although this should be confirmed by empirical research); the information is however available and artificial vision can conceivably find an use to it.

3.5 What does vision know about shadows?

Patrick Cavanagh (Paris Descartes University, FR)

License  Creative Commons BY 3.0 Unported license
© Patrick Cavanagh

How does the visual system identify and use shadow information in a scene? Shadow edges can irretrievably mislead the interpretation of a scene if they are taken to be object contours rather than shadow contours. So they need to be identified reliably and rapidly. Aside from their nuisance value, shadows also provide useful information about relative locations of objects. So how does vision identify shadows? It cannot be based on their shape as this is determined by several factors simultaneously: the direction of the light source, the shape of the object casting the shadow, and the surface relief on which it falls, as well as the relative positions of the light source, object, and receiving surface. In most cases this problem is intractable and clearly the visual system must have some set set of rules for rapidly identifying a shadow based on simple local properties. We will look at first the rules that artists have discovered in depicting convincing shadows and the rules of physics that artists know they can break. Then we will look at experimental evidence concerning the rules for shadows.

3.6 Shadows on the Moon and the Sun by Cigoli and Galilei: The Copernican planetarium inside Paolina's Chapel of Santa Maria Maggiore.

Marcos Cesar Danhoni Neves (State University of Maringá, BR)

License  Creative Commons BY 3.0 Unported license
© Marcos Cesar Danhoni Neves

Joint work of Danhoni Neves, Marcos Cesar; da Silva, Josie Agatha Parrilha

The Sidereus nuncius of Galileo Galilei irrefutably confirms the great revolution of the new astronomy initiated by Copernicus, Tycho, Kepler, Digges and Bruno. The presence of craters on the lunar surface showed, almost unequivocally, a marriage between heaven and earth, destroying the Aristotelian conception of the universe. However, this revolutionary idea of a Copernican universe was conquered not only by the acceptance of a new instrument, the telescope but, above all, by the construction frames of reference, the laws of linear and anamorphic perspective, and, especially, by the intense relationship between art-science, as can be seen by the friendship between Cigoli, Coccapani and Passignano with Galileo Galilei, on direct observation of the sun and its representations, including the famous moon cratered under the feet of the Madonna Assunta in the Paolina Chapel of Santa Maria Maggiore in Rome. But the shadow of the Inquisition condemns Galileo and erase the craters of the Madonna's Paolina, making it "Aristotelian", leading, also (as for the the nature of sunspots) to an arid intellectual debate, refusing the nature revealed by the perspective. However, a contemporary reinterpretation, and the rediscovery of the Cigoli's craters in twentieth century, allow a reconstruction of the new Galilean. This reconstruction allows us to discover a kind of a secret code inserted in the Cigoli's fresco. So, after the Church's censorship is removed, it is unveiled inside the Paolina dome a true Copernican planetarium, pushing back the shadows of the ecclesiastical censure since the condemnation of Galileo Galilei.

3.7 Why does computer vision find shadows so problematic?

Hannah M. Dee (Aberystwyth University, GB)

License © Creative Commons BY 3.0 Unported license
© Hannah M. Dee

Joint work of Dee, Hannah; Santos, Paulo E.

Main reference H. Dee, P. E. Santos, “The perception and content of cast shadows: an interdisciplinary review,”
Spatial Cognition and Computation, 11(3):226–253, 2011.

URL <http://dx.doi.org/10.1080/13875868.2011.565396>

This talk provides a brief review of the state of the art in shadow detection within computer vision. This considers image acquisition issues (single image, video from a static camera, video from a moving camera, video from a robot...), image processing issues (whether to work in the colour or texture domain), and issues of level of modelling (do we need to know the shape of the shadow caster? do we need to know where the light source is?). The talk will go on to emphasise issues and approaches for the evaluation of shadow detection algorithms, and will also show some datasets for shadow detection from a moving platform (wheeled, humanoid and UAV) which can be shared with workshop participants. Much progress has been made, particularly in the realm of shadow detection from a static camera in situations where statistical background models can be derived. However there remain many unsolved problems, and the issue of robotic shadow detection is largely unaddressed.

3.8 Shadows, Shape, Medialness

Frederic Fol Leymarie (University of London/Goldsmiths, GB)

License © Creative Commons BY 3.0 Unported license
© Frederic Fol Leymarie

We explore the shape of shadows in static and dynamic situations, via a representation based on medialness. Medialness is an extension of the classic medial graphs of Harry Blum and others. An image with contours, possibly fragmented, incomplete, is mapped to a medialness field under a gauge figure in the form of an annulus with width epsilon (a parameter alike a scaling factor). Ridge features extracted from this medialness field, such as peaks, saddles, ridge ends, provide a representation for the shape of the trace of objects, such as shadows.

We illustrate the retrieval of medial “hot” spots and significant convexities and concavities of shadow borders for static scenes as well as in videos (such as when observing a shadow theatre performance). The approach is inspired by recent results in perception, vision science and computational models. Differences between a shadow caster (3D) and its shadow (2D) are illustrated (e.g. with shadowgraphs). A careful study of the relation between the 3D nature of a physical object and the 2D nature of associated shadows remains an open and interesting avenue to explore, in static and dynamic situations.

3.9 Shadow and friends illuminate space

Christian Freksa (Universität Bremen, DE)

License © Creative Commons BY 3.0 Unported license
© Christian Freksa

Joint work of Freksa, Christian; Schultheis, Holger

Main reference C. Freksa, H. Schultheis, “Three Ways of Using Space,” in: D. R. Montello, K. E. Grossner, D. G. Janelle (eds.), “Space in mind: Concepts for spatial learning and education,” ISBN: 9780262028295, MIT Press 2014, pp. 31–48.

URL <https://mitpress.mit.edu/books/space-mind>

Cognitively relevant spatial parameters such as orientation, distance, and topological relationships are geometrically and physically determined by relative positions and orientations of objects in three-dimensional coordinate systems. Physically induced features such as visual illumination, occlusion, shadowing, reflection, mirroring; sound emanation and echo; surface texture and shape increase the complexity of sensory input to be processed by cognitive systems compared to scenes that are free of such features. However, as the geometric structures of these features are severely constrained by the structure of space and the physical laws of optics, acoustics, and haptics, respectively, they convey rich information about the space: they modulate the perceptibility of spatial configurations and environments and substantially enhance our abilities to disambiguate and make sense of perceptual stimulus patterns. Features such as shadows have been treated as noise that needs to be removed before spatial structures can be recognized; in contrast, we are interested in identifying ways in which these features can support recognition of spatial structures and spatial problem solving on the level of perceptual representation.

3.10 Shape-from-shadow polarity applies to 11 borders – 8 real, 3 pictorial

John Kennedy (University of Toronto, CA)

License © Creative Commons BY 3.0 Unported license
© John Kennedy

A. Ecological optics offers spectral and luminance borders. These can be static, and projected to a single vantage point. These are monocular borders. Optics also provides borders given by two vantage points. These are binocular (stereovision) and kinetic (accretion and deletion) borders. The total is $2 \times 2 \times 2 = 8$. These are real borders, I suggest. However, besides the 8 real borders, in outline pictures there are an additional 3 kinds of borders. Outline drawings have continuous lines, dotted lines and subjective-contour lines. Can all 11 trigger shape-from-shadow perception? Consider an observation of shadows and stains by Hering, and Cavanagh & Leclerc’s polarity hypothesis. For all 11 borders, shape-from-shadow would need low luminance on the shadow side, higher luminance on the illuminated side.

B. Lines can have two borders, one with correct polarity and one with reversed polarity. Countermanding the correct polarity, the reversed polarity border can block shape-from-shadow perception.

C. Shape-from-shadow vision should be successful even if the shadow borders have moving fringes (a demo from Bai).

D. P.S. 1. In a picture from Wnuczko, shadows cast on a plain and converging to a point on the horizon provide information that they are parallel. Does perception erroneously take the shadows as actually converging?

E. P.S. 2. In a picture from Hammad, shadows cast from a cube offer depicted angles and depicted angles. Are the depicting angles seen correctly, or biased towards the depicted angles?

F. P.S. 3. In a lunar eclipse, shadows cast on the moon by the Earth are information for the size of the Earth, as Kennedy and Casati have noted. They also indicate the size and distance of the sun. Can vision detect the sun’s size and distance?

3.11 Do Shadows Make Surfaces look Dark?

John O’Dea (University of Tokyo, JP)

License  Creative Commons BY 3.0 Unported license
© John O’Dea

Does a shadowed white area have a “appearance” like a non-shadowed grey area? Many say: In one sense “yes”, and in a other sense “no”. The “yes” sense is a problem, since it suggests that this aspect of perceptual experience is not representing things, or not representing straight- forwardly. The obvious solution is that we simply see the way things are illuminated in addition to the quality of the surface. In this talk, I focus on a particular way of modelling the experiential connection between surface colour and illumination quality: a multi-dimensional quality space. I am interested in one particular problem – that illumination and surface colour seem to be both integral and separable dimensions of the space.

3.12 Leveraging the Information in the Shadows of Synthetic Aperture Radar

Ann Marie Raynal (Sandia National Labs – Albuquerque, US)

License  Creative Commons BY 3.0 Unported license
© Ann Marie Raynal

Radar sensors have existed for nearly a century. However, radar data is not as intuitive as optical data for humans to interpret and glean important intelligence, surveillance, and reconnaissance information without extensive training or significant algorithm development and data post-processing to distill critical information. Well-designed radars, tools for their operation, training materials, and sensor data presentation or exploitation, can make immense impact on a mission’s execution by radar operators and analysts, whom are often overtasked and overwhelmed despite being incredibly capable. Radar shadows are the most optical-like features for human interpretation of radar data. Research from cross-disciplinary fields suggests shadows are innately preferred by our visual perception system for the interpretation of our spatial environment over other qualities. Preliminary research has shown shadows in radar imagery to be useful in determining stationary or moving target characteristics such as orientation, height, 3D shape, and location, which offer 3D imaging, detection, location, tracking, and identification possibilities. Unfortunately, shadows are inherently noisy and can disappear due to surrounding signals of even slightly higher energy that can overwhelm the

shadow. Furthermore, foibles of the human visual perception system can lead to erroneous assertions about objects and their spatial qualities. Shadow characteristics depend on many factors of the radar, platform, object, and scene. Algorithm development and requirements are needed for more automated analysis of shadows to overcome human flaws and shadow degradation. The utility of radar shadows and how best to leverage their information for radar applications is investigated in this talk.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

- 1 B. T. Backus and I. Oruç, "Illusory motion from change over time in the response to contrast and luminance," *Journal of Vision*, 5(11), December 30, 2005.
- 2 G. Mather, "Luminance change generates apparent movement: implications for models of directional specificity in the human visual system," *Vision Res.*, 24(10):1399–1405, 1984.
- 3 H. M. Dee and P. E. Santos, "The perception and content of cast shadows: an interdisciplinary review," *Spatial Cognition & Computation: An Interdisciplinary Journal*, 11(3):226–253, 2011.
- 4 S. Taya and K. Miura, "Cast shadow can modulate the judged final position of a moving target," *The Physiognomic Society, Attention, Perception, & Psychophysics*, 72(7):1930–1937, 2010.
- 5 W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*, Artech House, Boston, 1995.
- 6 F. M. Dickey and A. W. Doerry, "Recovering shape from shadows in synthetic aperture radar imagery," *Proc. SPIE*, 6947(694707), 2008.
- 7 J. Groen, R. E. Hansen, H. J. Callow, J. C. Sabel, and T. O. Sabo, "Shadow enhancement in synthetic aperture sonar using fixed focusing," *IEEE Journal of Oceanic Engineering*, 34(3):269–284, July 2009.
- 8 H. J. Callow, J. Groen, R. E. Hansen, and T. Sparr, "Shadow enhancement in SAR imagery," 2007 IET International Conference on Radar Systems, pp. 1–5, Oct. 15–18, 2007.
- 9 M. Jahangir, "Moving target detection for synthetic aperture radar via shadow detection," 2007 IET International Conference on Radar Systems, pp. 1–5, 15–18 Oct. 2007.
- 10 L. M. Novak, "Optimal target designation techniques," *IEEE Transactions on Aerospace and Electronic Systems*, 17(5):676–684, Sept. 1981.
- 11 F. T. Ulaby and M. C. Dobson, *Radar Scattering Statistics for Terrain*, Artech House, Inc., Norwood, MA, 1989.
- 12 H. A. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Transactions on Geoscience and Remote Sensing*, 30(5):950–959, Sep 1992.
- 13 D. L. Bickel, "SAR Image Effects on Coherence and Coherence Estimation," Sandia National Laboratories Report, Albuquerque, NM, SAND2014-0369, January 2014.
- 14 A. M. Raynal, D. L. Bickel, and A. W. Doerry, "Stationary and Moving Target Shadow Characteristics in Synthetic Aperture Radar," *Proc. SPIE*, 9077(90771B-1), 2014.
- 15 A. M. Raynal, J. Miller, E. Bishop, V. Horndt, and A. W. Doerry, "Shadow Probability of Detection and False Alarm for Median-Filtered SAR Imagery," Sandia National Laboratories Report, Albuquerque, NM, SAND2014-4877, June 2014.
- 16 J. Miller, E. Bishop, V. Horndt, A. W. Doerry, and A. M. Raynal, "Impact of Ground Mover Motion and Windowing on Stationary and Moving Shadows in Synthetic Aperture Radar Imagery," *Proc. SPIE*, April 2015.
- 17 C. K. Peterson, P. Murphy, and P. Rodriguez, "Target Classification in Synthetic Aperture Radar Using Map-Seeking Circuit Technology," *Proc. SPIE*, 8051(13), 2011.

3.13 Shadows in Robotics and Artificial Intelligence

Paulo E. Santos (Centro Universitario da FEI – São Paulo, BR)

License © Creative Commons BY 3.0 Unported license
© Paulo E. Santos

Joint work of Santos, Paulo E.; Dee, Hannah M.; Fenelon, Valquiria

Main reference H. Dee, P. E Santos, “The Perception and Content of Cast Shadows: An Interdisciplinary Review,”
Spatial Cognition and Computation, 11(3):226–253. 2011.

URL <http://dx.doi.org/10.1080/13875868.2011.565396>

Recent research in psychology suggests that the human perceptual system gives preferential treatment to information from shadows when inferring motion in depth and perceiving 3D scene layout. Much work in computer vision and robotics, however, starts from the premise that shadows are sources of noise rather than information. Indeed, due to high contrast in the image, shadows are among the most salient visual items, and a source of distraction from objects and locations.

The purpose of this talk is to bring together contributing ideas from the various disciplines involved in the investigation of the knowledge content in cast shadows in artificial systems. This talk starts with a short survey of the use of the information content of shadows in the fields of AI and Robotics and finishes with a proposal of a probabilistic algorithm for robot self-localisation that is based on a topological map constructed from the observation of cast shadows. Distinct locations on the map are defined by means of a classical formalism for qualitative spatial reasoning.

3.14 Grasping Objectified Shadows

Carl Schultz (Universität Münster, DE)

License © Creative Commons BY 3.0 Unported license
© Carl Schultz

Architects have long recognised that it is useful to treat shadows as real, physical, “first-class” objects, at a similar ontological level as objects with a material extension (furniture, walls, and so on). It is useful in the sense that the spatial patterns of light and shadow that people perceive have a powerful influence on their subjective impressions of the environment; as architects are ultimately concerned with designing experiences, identifying salient light-shadow patterns produced by the interaction of shadow objects with the environment is of central importance. I am interested in how we can formally model these rules about experience and patterns within artificial intelligence representation and reasoning frameworks. The aim is to provide cognitively- driven design analysis support in a high-level, qualitative and declarative manner.

3.15 Shadow Messages in the Arts

William Sharpe (Columbia University, US)

License © Creative Commons BY 3.0 Unported license
© William Sharpe

In the arts, obvious “messages” are frowned upon; didactic work has been largely out of favor for over a century. And looking back over earlier periods, scholars, critics, and artists

themselves generally prefer works that seem to demand interpretation, or that seem to propose a number of plausible “readings”. For artist and writers, then, the challenge is to make their shadow “messages” intriguing, without saying too much or too little. Sometimes the desire is to minimize shadow “noise”, but other times the aim is to increase that “noise” into the central statement of the artwork. In a forthcoming book I argue that in literature and the visual arts there are basically four kinds of shadow, which can be distinguished by their use (on the creative end) and by their reception (on the audience’s end). Painters, writers, photographers, and filmmakers employ a spectrum of realistic shadows, from nearly imperceptible shadows to strongly defined dramatic ones, to provide information about the light, volume, and location of objects and people within a given scene. Often they seek to accomplish this without distracting the reader/viewer’s attention from the main features of the narrative or scene. But they may also desire to activate these shadow functions (light, location, volume) via the use of shadows that add special emotional and semantic features—moody, menacing, vibrant, grotesque shadows. We regularly meet with what I call “Look Elsewhere” and “Vital” shadows, shadows that either redirect the attention “elsewhere” to their casting objects, or shadows that seem expressively to call attention to themselves (what Roberto Casati has called “reflexive shadows”). But sometimes even further semantic information is added, to the detriment or transcending of informational aims. Separating shadows from their casters, the shadow-makers foreground human emotional investment in shadows. “Completing” shadows reveal the void that is felt when objects or people are bereft of shadows that signal full participation in life; and “Independent” shadows, becoming permanently detached from their “owners”, take on an autonomous existence as dark impulses and rebellious figures seeking to revolutionize the artistic landscape. In artistic expression, the more “noise” that is created by a noticeable shadow, the stronger is the message of the shadow. But what that message is depends on our culturally determined understanding of shadows, particularly what we perceive as their troublingly wayward behavior. Information of a concrete sort may be hard to come by; like shadows, shadow- messages in word or image are a challenge to grasp. At the least, however, we can sort out what kind of shadow we are encountering, and (roughly) what the range of meanings is that can be assigned to it.

3.16 Declarative Methods for Cognitive Vision

Jakob Suchan (Universität Bremen, DE)

License © Creative Commons BY 3.0 Unported license
© Jakob Suchan

Main reference J. Suchan, M. Bhatt, P.E. Santos, “Perceptual Narratives of Space and Motion for Semantic Interpretation of Visual Data,” in Proc. of the Computer Vision – ECCV 2014 Workshops – Part II, LNCS, Vol. 8926, pp. 339–354, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-16181-5_24

The cognitive interpretation of interactions of humans in their everyday surroundings requires systems to perceive their operational environment, recognise activities performed, and reason about the observed information. This includes the detection of low-level perceptual information about the individuals and objects in the scene, motion, shape, relative configuration, and the computational capability to interpret these in terms of high-level human interactions performed towards the goal of an activity.

Our research focuses on developing declarative models of human interactions that are grounded in their low-level / quantitative perceptual input. The aim is to allow systems to perform integrated reasoning about space, actions, and change at a commonsense level. The

implemented model is such that primitives of the theory, e.g., pertaining to space and motion, are available as first-class objects with deep semantics suited for inference and query. As applications for this research, we focus on the cognitive interpretation of human interactions in areas such as smart environments, human behaviour in indoor space (e.g., for wayfinding analysis), and cognitive studies pertaining to the embodied perception and interpretation of films.

3.17 The systematic introduction of Chiaroscuro in 15th century Florence and the symbolic shadow in Sieneese Painting and Un-naturalistic painting

Koichi Toyama (Keio University, JP)

License  Creative Commons BY 3.0 Unported license
© Koichi Toyama

Illusionism in painting using cast shadows was first introduced by Masaccio in the mid-1420s at the Santa Maria del Carmine, Florence. There, the painter depicted naturalistic cast shadows in scenes such as the The Tribute Money, The Expulsion from the Garden of Eden, and St. Peter Healing the Sick with His Shadow, as if the shadows were projected by light penetrating from the rear window of the Brancacci Chapel. It is clear that this system of shadow was added to the system of shading introduced by Giotto at the Scrovegni Chapel, Padua, almost one century before. I will take several examples from contemporary Sieneese religious paintings in which painters such as Sassetta, Giovanni di Paolo, and Pietro di Giovanni d’Ambrogio depicted cast shadows only once or twice during their lifetimes, supposedly for specific symbolic reasons. Then I will discuss examples of incongruous shadows found even in “realistic” Florentine painting.

3.18 Un-naturalistic painting and the lack of shadow: History of shadow in 18th-19th century Japanese paintings and woodblock prints

Koichi Toyama (Keio University, JP)

License  Creative Commons BY 3.0 Unported license
© Koichi Toyama

Historically speaking, no culture was able to depict cast shadows systematically without contact with Western painting from after the Renaissance. Japan was no exception. During the national isolation of the Edo period (1603-1868), only after the import of Dutch books was permitted in the mid-18th century did a few Japanese intellectuals and artists attempt to learn Western illusionistic techniques such as linear perspective and chiaroscuro. Afterwards, certain painters who created designs for woodblock prints attempted to adopt cast shadows only for specific scenes. I would like to illustrate the reception of chiaroscuro in Japan using the works of Hokusai and Hiroshige (Night View of Saruwaka-cho).

3.19 Shadow Play

Barbara Tversky (Stanford University, US)

License  Creative Commons BY 3.0 Unported license
© Barbara Tversky

Shadow and shadows have many meanings, and are used deliberately by speakers, photographers, filmmakers, cartoonists, and artists to create meaning. Can these creations inform perception? I present a loose collection of uses of shadows in language, film, photography, and art that illustrate some of the subtle meanings shadows convey.

3.20 Perception of Shadow and Shading in Paintings and Photographs

Maarten Wijntjes (TU Delft, NL)

License  Creative Commons BY 3.0 Unported license
© Maarten Wijntjes

Our study starts with the observation that painters often do not comply to the rules of linear perspective when it comes to cast shadows. Cast shadows are relatively rare in paintings (although they seem to receive more popularity from the 19th century onwards), especially cast shadows that go beyond the a blurry, formless blob. Canaletto (1697–1768) was certainly not avoiding cast shadows in his “vedute” of Venice. Furthermore, it appears that he was aware of potential problems with perspective since he (and many other painters) often orients the sunlight parallel to the projection plane. Therefore, the cast shadows are also parallel to the projection plane and do not have to converge to a vanishing point in the picture plane. In some cases Canaletto does use an oblique illumination and the problems arise quite evidently: whereas the architectural content complies perfectly to the rules of perspective, the cast shadows do not vanish in a single vanishing point.

This observation made us look a bit closer at the shadow casting persons in these paintings. It occurred to us that in quite a number of cases, the shading direction did not match the shadow direction. Whereas shadows often indicated that the light came from exactly left or right, the shading suggested more oblique angles, often more from the front. When the light comes somewhat from the front, the humans are better visible and may look better. Therefore, we hypothesised that Canaletto (and possibly other painters) uses two different motivations for the rendering of shading and shadowing: aesthetical and practical.

This finding made us question how sensitive humans are for possible discrepancies between shading and shadowing. We performed a light direction estimation experiment in which small fragments of Canaletto paintings are presented flanked by an interactive light probe. The light direction on this probe could be adjusted to match the illumination on the fragment. Three conditions were used: shadow only (persons were black silhouettes), shading only, and combined. We found that for the shadow only and shading only conditions light direction is estimated markedly different. In the combined condition we found that some observers rely purely on the shadow cue, whereas others combine the cues. In a follow up we also used photographs. This data is still being analysed but should be ready at the Dagstuhl seminar.

Participants

- Juliano Beraldo
University São Paulo, BR
- Mehul Bhatt
Universität Bremen, DE
- Francesca Bizzarri
Associazione Ca' Luogo D'Arte –
Gattatico, IT
- Roberto Casati
ENS – Paris, FR
- Patrick Cavanagh
Paris Descartes University, FR
- Marcos Cesar Danhoni Neves
State University of Maringás, BR
- Hannah Dee
Aberystwyth University, GB
- Valquiria Fenelon Pereina
University Center of FEI –
São Paulo, BR
- Frederic Fol Leymarie
Goldsmiths – University of
London, GB
- Christian Freksa
Universität Bremen, DE
- John Kennedy
University of Toronto, CA
- John O'Dea
University of Tokyo, JP
- Ann Marie Raynal
Sandia National Labs –
Albuquerque, US
- Paulo E. Santos
University Center of FEI –
São Paulo, BR
- Carl Schultz
Universität Münster, DE
- William Sharpe
Columbia University, US
- Jakob Suchan
Universität Bremen, DE
- Koichi Toyama
Keio University, JP
- Barbara Tversky
Stanford University, US
- Maarten Wijntjes
TU Delft, NL



Report from Dagstuhl Seminar 15201

Cross-Lingual Cross-Media Content Linking: Annotations and Joint Representations

Edited by

Alexander G. Hauptmann¹, James Hodson², Juanzi Li³,
Nicu Sebe⁴, and Achim Rettinger⁵

1 Carnegie Mellon University, US, alex@cs.cmu.edu

2 Bloomberg – New York, US, jhodson2@bloomberg.net

3 Tsinghua University – Beijing, CN, lijuanzi@tsinghua.edu.cn

4 University of Trento, IT, sebe@disi.unitn.it

5 KIT – Karlsruher Institut für Technologie, DE, rettinger@kit.edu

Abstract

Dagstuhl Seminar 15201 was conducted on “Cross-Lingual Cross-Media Content Linking: Annotations and Joint Representations”. Participants from around the world participated in the seminar and presented state-of-the-art and ongoing research related to the seminar topic. An executive summary of the seminar, abstracts of the talks from participants and working group discussions are presented in the forthcoming sections.

Seminar May 10–13, 2015 – <http://www.dagstuhl.de/15201>

1998 ACM Subject Classification I.2.7 Natural Language Processing, I.2.10 Vision and Scene Understanding, H.3.3 Information Search and Retrieval, I.2.4 Knowledge Representation Formalisms and Methods

Keywords and phrases Cross-lingual, Cross-media, Cross-modal, Natural language processing, Computer vision, Multimedia, Knowledge representation, Machine learning, Information extraction, Information retrieval

Digital Object Identifier 10.4230/DagRep.5.5.43

Edited in cooperation with Aditya Mogadala

1 Executive Summary

Alexander G. Hauptmann

James Hodson

Juanzi Li

Nicu Sebe

Achim Rettinger

License  Creative Commons BY 3.0 Unported license

© Alexander G. Hauptmann, James Hodson, Juanzi Li, Nicu Sebe, and Achim Rettinger

Different types of content belonging to multiple modalities (text, audio, video) and languages are generated from various sources. These sources either broadcast information on channels like TV and News or allow collaboration in social media forums. Often multiple sources are consumed in parallel. For example, users watching TV tweeting their opinions about a show. This kind of consumption throw new challenges and require innovation in the approaches to enhance content search and recommendations.



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

Cross-Lingual Cross-Media Content Linking: Annotations and Joint Representations, *Dagstuhl Reports*, Vol. 5, Issue 5, pp. 43–56

Editors: Alexander G. Hauptmann, James Hodson, Juanzi Li, Nicu Sebe, and Achim Rettinger



DAGSTUHL
REPORTS

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

Currently, most of search and content based recommendations are limited to monolingual text. To find semantic similar content across different languages and modalities, considerable research contributions are required from various computer science communities working on natural language processing, computer vision and knowledge representation. Despite success in individual research areas, cross-lingual or cross-media content retrieval has remained an unsolved research issue.

To tackle this research challenge, a common platform is provided in this seminar for researchers working on different disciplines to collaborate and identify approaches to find similar content across languages and modalities. After the group discussions between seminar participants, two possible solutions are taken into consideration:

1. *Building a joint space from heterogeneous data generated from different modalities to generate missing or to retrieve modalities. This is achieved through aligned media collections (like parallel text corpora). Now to find cross-media cross-lingual relatedness of the content mapped to a joint latent space, similarity measures can be used.*
2. *Another way is to build a shared conceptual space using knowledge bases(KB) like DBpedia etc for semantic annotation of concepts or events shared across modalities and languages. Entities are expressed in any channel, media type or language can be mapped to a concept space in KB. Identifying a commonality between annotations can be used to find cross-media cross-lingual relatedness.*

Thus, implementing these solutions require a joint effort across research disciplines to relate the representations and to use them for linking languages and modalities. This seminar also aimed to build datasets that can be used as standard test bed and benchmark for cross-lingual cross-media content linking. Also, seminar was very well received by all participants. There was a common agreement that the areas of text, vision and knowledge graph should work more closely together and that each discipline would benefit from the other. The participants agreed to continue to work on two cross-modal challenges and discuss progress and future steps in a follow-up meeting in September at Berlin .

2 Table of Contents

Executive Summary

Alexander G. Hauptmann, James Hodson, Juanzi Li, Nicu Sebe, and Achim Rettinger 43

Overview of Talks

NLU for Colloquial Text and Speech-to-text <i>Xavier Carreras</i>	46
Data Analytics over Multiple Content Types <i>John Davies</i>	47
Cross-Domain Cue Switching <i>Tiansi Dong</i>	47
Cross-Lingual Document Similarity and Event Tracking <i>Blaz Fortuna</i>	48
NELL as a Knowledge Graph building tool <i>Estevam R. Hruschka</i>	48
Multi Lingual Knowledge Graph <i>Juanzi Li</i>	49
Extracting aggregated knowledge from cross-lingual news <i>Dunja Mladenic</i>	49
Multimodal Learning <i>Aditya Mogadala</i>	50
Bloomberg Named Entity Disambiguation <i>Stefano Pacifico</i>	50
Machine Learning, Image Annotation and Computer Vision <i>Alan Smeaton</i>	51
Relational Machine Learning for Knowledge Graphs <i>Volker Tresp</i>	51
Automatic extraction of ontology lexica in multiple languages <i>Christina Unger</i>	52
Learning Knowledge Graphs from Images and Text <i>Lexing Xie</i>	52

Working Groups

Working Group II: State-of-the-art Text and Knowledge Graphs <i>Estevam R. Hruschka</i>	53
Working Group III: Visual Information and Knowledge Graphs <i>Dubravko Culibrk</i>	53
Working Group IV: Representation Learning <i>Aditya Mogadala</i>	54

Open Problems 55

Participants 56

3 Overview of Talks

3.1 NLU for Colloquial Text and Speech-to-text

Xavier Carreras (Xerox Research Centre Europe – Grenoble, FR)

License  Creative Commons BY 3.0 Unported license
© Xavier Carreras

State-of-the-art approaches to Natural Language Understanding (NLU) are based on supervised statistical techniques, and thus rely on the availability of treebanks, i.e. textual collections annotated with linguistic structure. Most available treebanks today annotate newswire articles, which are characterized by being edited text written by professionals. In contrast, many NLU applications deal with text generated by non-professional writers, which in most cases is produced spontaneously in conversations. This is the case for most of web data, emails, dialogue systems, and social media chatter. In these type of textual data, it is common to find spelling mistakes and ungrammatical constructs. In addition, the distribution of topics and words will in most cases differ significantly from newswire data. All these facts pose difficult challenges for NLU on colloquial text. Similarly, in applications dealing with speech, the automatic conversion from speech to text results in noisy textual data which significantly differs from the mistake-free text we find in treebanks.

In this talk I will review recent work in the state-of-the-art for NLU on colloquial text and speech data. I will describe work in three directions. The first is applying domain adaptation techniques, where I will review the main conclusions of the recent “Parsing the Web” challenge [1], and I will describe the architecture of the top-performing system (Le Roux et al 2012).

Then I will describe work for linguistic analysis of Twitter by Gimpel et al. [2], Owoputi et al. [3], and Kong et al. [4]. In these approaches, tweets are seen so dramatically different than standard newswire data that authors choose to redefine the classic annotation standards and reannotate data. In other words, authors start a new task from scratch, attempting to come up with linguistic annotations that reflect the nature of tweets, as opposed to trying to describe tweets as if they were news articles. Some challenges here are how to deal with spontaneous variations of tokens, and how to take advantage of the large and rich resources we have for standard domains.

Finally I will describe some approaches for Spoken Language Understanding. To cope with the noisy input, one approach is to directly represent the confusion word lattice with various features that capture the uncertainty.

Overall, these techniques show a degree of improvements. But the problem of NLU on colloquial text and speech-to-text remains a challenging, difficult and important open problem.

References

- 1 Petrov, S., and McDonald, R.: Overview of the 2012 shared task on parsing the web. In Notes of the First Workshop on Syntactic Analysis of Non-Canonical Language (SANCL) (2012)
- 2 Gimpel, K., Schneider, N., O’Connor, B., Das, D., Mills, D., Eisenstein, J., and Smith, N. A. Part-of-speech tagging for twitter: Annotation, features, and experiments. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics. (2011)

- 3 Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., and Smith, N. A. Improved part-of-speech tagging for online conversational text with word clusters. Association for Computational Linguistics. (2013)
- 4 Kong, L., Schneider, N., Swayamdipta, S., Bhatia, A., Dyer, C., and Smith, N. A. A dependency parser for tweets. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. (2014)

3.2 Data Analytics over Multiple Content Types

John Davies (BT Research – Ipswich, GB)

License  Creative Commons BY 3.0 Unported license
© John Davies

We present 3 use cases of the use of data analytics over multiple content types. In the first case, structured data and unstructured data (text documents) relating to an organization's sales performance are analyzed. Named entity recognition is used to identify people, products, companies, locations and key-phrases in the textual data and associated with an ontology using semantic annotation. The structured data is held in a traditional RDB and a dashboard is developed allowing queries over both data types allowing the data to be analyzed for improved sales management information in ways previously only possible via a time-consuming manual process. In the second use case, we combine sensor data with social media data. We exemplify in the transport domain, where tweets about traffic are filtered and their location is extracted automatically from the tweet content. This can then be combined with roadside sensor data giving a combined view of both numeric values such as traffic speed and density and Twitter users observations about the same road segment. This is useful for highways authorities where valuable information relating to traffic incidents is often to be found in social media content. Finally, we discuss the value of combining video content with social media data and specifically of aligning events described or shown in both media. Here the example is sport, where incidents in sports games could be detected by video recognition (and/or textual analysis of subtitles) and linked to social media users' reactions to the same event. Media companies can thereby gauge reaction to certain events and better understand their customers.

3.3 Cross-Domain Cue Switching

Tiansi Dong (Universität Bonn, DE)

License  Creative Commons BY 3.0 Unported license
© Tiansi Dong

Descriptions of the same entity may capture different meaning aspects, if we choose different media. Cue, as used mainly in psychology, refers to any piece of information between descriptions and meaning aspects. To transform a description from the source media into a form in the target media, we shall first retrieve the meaning aspect of the description in the source domain, transform it into the meaning aspect of the target media, and deliver descriptions of the target media. Three examples in natural language translation are presented. The first example is to translate “white as snow” into the native language of Benin, where there is no word for “snow”; the second example is to translate “you are my heart into

Indonesian”, where hearts are regarded less important than livers; the third example is to translate the description “the western table” into “the table on my left”, which can only be achieved through a spatial transformation. Such spatial transform can be achieved by understanding orientation relations as distance comparison relations. We further show how orientation relations shall be understood as distance comparison relations in general, and how distance comparison and distance relations can be defined in the connection relation. As spatial domain is the first domain human babies encounter and understand, this domain is used as the reference domain for the cognition of other domains. Thus, spatial domain is the base domain for cross-domain cue switching. Our current research work on German-Chinese cue-switching translation is outlined. Cues used in other domains are listed.

3.4 Cross-Lingual Document Similarity and Event Tracking

Blaz Fortuna (Ghent University, BE)

License © Creative Commons BY 3.0 Unported license

© Blaz Fortuna

Joint work of Rupnik, Jan; Muhic, Andrej; Leban, Gregor; Skraba, Primoz; Fortuna, Blaz; Grobelnik, Marko;

URL <http://xling.ijs.si/>

In this work, we address the problem of tracking and events in a large multilingual stream. We consider a particular aspect of this problem, namely how to link collections of articles in different languages which refer to the same event.

Given a multi-lingual stream and clusters of articles from each language, we propose a method for cross-lingual document similarity based on Wikipedia, which enables us to compute the similarity of any two articles regardless of language. The approach learns an representations of documents which were valid over multiple languages. The representations could be interpreted as multi-lingual topics, which were then used as proxies to compute cross-lingual similarities between documents. To learn the representations, we use Wikipedia as a training corpus. Significantly, we do not only consider the major or hub languages such as English, German, French, etc. which have significant overlap in article coverage, but also smaller languages (in terms of number of Wikipedia articles) such as Slovenian and Hindi, which may have a negligible overlap. The proposed method can scale to 100 languages and can match articles from languages with little or no direct overlap in the training data.

3.5 NELL as a Knowledge Graph building tool

Estevam R. Hruschka (University of Sao Carlos, BR)

License © Creative Commons BY 3.0 Unported license

© Estevam R. Hruschka

Joint work of Mitchell, Tom M.; Cohen, William W.; Hruschka, Estevam R.

Main reference T. M. Mitchell, W. W. Cohen, E. R. Hruschka Jr., P. P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi Mishra, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. A. Platanios, A. Ritter, M. Samadi, B. Settles, R. C. Wang, D. T. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, J. Welling, “Never-Ending Learning,” in Proc. of the 29th AAAI Conf. on Artificial Intelligence (AAAI’15), pp. >2302–2310, AAAI Press, 2015.

URL <http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/10049>

Never-Ending Language Learner (NELL) is a computer system that runs 24/7, forever, learning to read the web. The system is designed to perform two basic tasks: i) extract (read) more facts from the web, and integrate these into its growing knowledge base of beliefs; and

ii) learn to read better than yesterday, enabling it to go back to the text it read yesterday, and today extract more facts, more accurately. This system has been running 24 hours/day for over four years now. The result so far is a collection of 90 million interconnected beliefs (e.g., `servedWith(coffee, applePie)`, `isA(applePie, bakedGood)`), that NELL is considering at different levels of confidence, along with hundreds of thousands of learned phrasings, morphological features, and web page structures that NELL uses to extract beliefs from the web.

3.6 Multi Lingual Knowledge Graph

Juanzi Li (Tsinghua University – Beijing, CN)

License  Creative Commons BY 3.0 Unported license
© Juanzi Li

Multilingual knowledge graph is the graph of entities and relationships in different languages. Multilingual knowledge graph, recognized as the bridges for information understanding across multiple languages, can enhance the linked data internationalization and globalization of knowledge sharing among different languages on the Web and, facilitate the cross-lingual language processing such as cross-lingual information retrieval, machine translation and question answering etc. In this talk, we summarize the existing monolingual and multilingual knowledge graphs, and the state of the art technologies including multilingual knowledge linking, multilingual knowledge building and cross lingual knowledge extraction. The talk is concluded with some identified challenging problems such us multilingual knowledge representation learning, multimedia and multilingual knowledge linking based on it.

3.7 Extracting aggregated knowledge from cross-lingual news

Dunja Mladenic (Jozef Stefan Institute – Ljubljana, SI)

License  Creative Commons BY 3.0 Unported license
© Dunja Mladenic

Joint work of Mladenic, Dunja; Marko Grobelnik; Blaz Fortuna; Gregor Leban; Blaz Noak; Jan Rupnik; Mitja Trampus; Andrej Muhic

Cross-lingual news analysis in general requires handling large amount of textual data across different languages. We propose combining Machine Learning and Natural Language Processing methods to enable extracting aggregated knowledge from cross-lingual news. In particular, we propose several lines of development involving research and development of prototype systems for news annotation in real-time, mining event patterns, identifying events across languages, detect diversity of reporting along several dimensions, rich exploratory visualizations of news events, interoperable data export.

We demonstrate the functioning on operational news monitoring and extracting knowledge that involves a large number of data streams in multiple languages. The following are publicly available related systems: Event Registry¹ for event detection and topic tracking;

¹ <http://eventregistry.org/>

DiversiNews² for news diversity explorer; NewsFeed³ for news and social media crawler; Enrycher⁴ for language and semantic annotation; XLing⁵ for cross-lingual document linking and categorization.

3.8 Multimodal Learning

Aditya Mogadala (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license
© Aditya Mogadala

The growth of multimedia content on the web raise diverse challenges. Over the decades various approaches are designed to support search, recommendations, analytics and advertising based on the textual content. But now due to the overwhelming availability of multimedia content require update in technologies to leverage multimedia information. Recent advancements made in machine learning to foster continuous representations of text and effectual object detection in videos and images provide new opportunities. In this aspect, leveraging data generated from videos, images and text to support various applications by finding cross-modal semantic similarity. In particular, to compare semantically similar content generated across media by jointly modeling two different modalities.

Modeling one or more modalities together can be helpful to generate missing modalities and retrieve cross-modal content. Results are overwhelming when used to jointly model images or videos along with captions using deep learning approaches like Recurrent neural networks and multimodal log-bilinear models. It also pave the path to extend textual information to multiple languages for supporting the growth of polylingual content on the web.

3.9 Bloomberg Named Entity Disambiguation

Stefano Pacifico (Bloomberg – New York, US)

License © Creative Commons BY 3.0 Unported license
© Stefano Pacifico

Joint work of Pacifico, Stefano; Bradesko, Luka; Starc, Janez
Main reference L. Bradesko, J. Starc, S. Pacifico, Stefano, “Isaac Bloomberg Meets Michael Bloomberg: Better Entity Disambiguation for the News,” in Proc. of 24th Int’l World Wide Web Conference (WWW’15) – Companion Volume, pp. 631–635, ACM, 2015.
URL <http://dx.doi.org/10.1145/2740908.2741711>

In this talk we present BNED, the Named Entity Disambiguation system developed and used at Bloomberg. In particular, we illustrate how we built a system that do not require the use of Wikipedia as a knowledge base or training corpus. We also present how we built features for disambiguation algorithms significative for the Bloomberg News corpus, and show results of both single-entity and joint-entity disambiguation into the Bloomberg proprietary knowledge base of people and companies.

² <http://aidemo.ijs.si/diversinews/>

³ <http://newsfeed.ijs.si/>

⁴ <http://enrycher.ijs.si/>

⁵ <http://xling.ijs.si/>

3.10 Machine Learning, Image Annotation and Computer Vision

Alan Smeaton (Dublin City University, IE)

License © Creative Commons BY 3.0 Unported license
© Alan Smeaton

This presentation covered an overview of the state-of-the-art in automatic detection of semantic concepts from visual media. The presentation started with an overview of the challenges in captioning or tagging or annotating visual media and then described how we use low-level image features – colours, textures, shapes, SIFT/SURF features – to index images so we can support look-alike visual similarity searching. This is useful in some applications but doesn't directly address the problem of describing an image's contents. We then moved on to present how the multimedia indexing field uses off-the-shelf machine learning to build classifiers, usually one at a time, and how the performance of these has evolved and improved over the last decade in the TRECVID benchmarking. We then looked at most recent work on image captioning using deep learning from groups in Stanford and in Google, as reported in the NYT, and finally ended the presentation with a showcase of building classifiers in real time, during a search, the advantages of this approach being that we don't have to know what people might want to search for in advance.

3.11 Relational Machine Learning for Knowledge Graphs

Volker Tresp (Siemens AG – München, DE)

License © Creative Commons BY 3.0 Unported license
© Volker Tresp

Joint work of Tresp, Volker; Maximilian Nickel; Denis Krompass; Xueyan Jiang

Main reference M. Nickel, K. Murphy, V. Tresp, E. Gabrilovich, "A Review of Relational Machine Learning for Knowledge Graphs: From Multi-Relational Link Prediction to Automated Knowledge Graph Construction," to appear as invited paper in the Proceedings of the IEEE; pre-print available as arXiv:1503.00759v3 [stat.ML].

URL <http://arxiv.org/abs/1503.00759v3>

Most successful applications of statistical machine learning focus on response learning or signal-reaction learning where an output is produced as a direct response to an input. An important feature is a quick response time, the basis for, e.g., real-time ad-placement on the Web, real-time address reading in postal automation, or a fast reaction to threats for a biological being. One might argue that knowledge about specific world entities and their relationships is necessary if the complexity of an agent's world increases, for example if an agent needs to function in a complex social community. As one is quite aware in the Semantic Web community, a natural representation of knowledge about entities and their relationships is a directed labeled graph where nodes represent entities and where a labeled link stands for a true fact. A number of successful graph-based knowledge representations, such as DBpedia, YAGO, or the Google Knowledge Graph, have recently been developed and are the basis of applications ranging from the support of search to the realization of question answering systems. Statistical machine learning can play an important role in knowledge graphs as well. By exploiting statistical relational patterns one can predict the likelihood of new facts, find entity clusters and determine if two entities refer to the same real world object. Furthermore, one can analyze new entities and map them to existing entities (recognition) and predict likely relations for the new entity. These learning tasks can elegantly be approached by first transforming the knowledge graph into a 3-way tensor where two of the modes represent the

entities in the domain and the third mode represents the relation type. Generalization is achieved by tensor factorization using, e.g., the RESCAL approach. A particular feature of RESCAL is that it exhibits collective learning where information can propagate in the knowledge graph to support a learning task. In the presentation the RESCAL approach will be introduced and applications of RESCAL to different learning and decision tasks will be presented.

3.12 Automatic extraction of ontology lexica in multiple languages

Christina Unger (Universität Bielefeld, DE)

License  Creative Commons BY 3.0 Unported license
© Christina Unger

Many applications that need to mediate between natural language and Semantic Web data, such as question answering and verbalization of ontologies or RDF datasets, require knowledge about how elements of the vocabulary are expressed in natural language. Moreover, in case a system is supposed to be multilingual, this knowledge is needed in multiple languages. In this talk, I present a model for capturing such lexical knowledge as well as a recent approach to automatically acquiring it, and outline the main limitations this approach still faces.

3.13 Learning Knowledge Graphs from Images and Text

Lexing Xie (Australian National University – Canberra, AU)

License  Creative Commons BY 3.0 Unported license
© Lexing Xie

Main reference L. Xie, H. Wang, “Learning Knowledge Bases for Text and Multimedia ,” ACM Multimedia 2014 Tutorial, 2014.

URL http://users.cecs.anu.edu.au/~xlx/proj/knowledge_mm14.html

Knowledge acquisition, representation, and reasoning have been one of the long-standing challenges in artificial intelligence and related application areas. Only in the past few years, massive amounts of structured and semi-structured data that directly or indirectly encode human knowledge became widely available, turning the knowledge representation problems into a computational grand challenge with feasible solutions in sight. The research and development on knowledge bases is becoming a lively fusion area among web information extraction, machine learning, databases and information retrieval, with knowledge over images and multimedia emerging as another new frontier of representation and acquisition. This tutorial aims to present a gentle overview of knowledge bases on text and multimedia, including representation, acquisition, and inference. I present a brief survey of work on learning words, entities and their relations from images and their accompanying words.

4 Working Groups

4.1 Working Group II: State-of-the-art Text and Knowledge Graphs

Estevam R. Hruschka (University of Sao Carlos, BR)

License  Creative Commons BY 3.0 Unported license
© Estevam R. Hruschka

One subgroup was focused on discussing and motivating all the participants to discuss the state-of-the-art bridging Text and Knowledge Graphs, and how Cross-Lingual, as well as Cross-Media can be explored to help in defining joint Representations. The discussions led to some very interesting issues, and most of those are based on the fact that it is very difficult to define an optimum representation (ontology) describing an ideal taxonomy that would allow:

- (i) Using the current state-of-the-art to generate more “useful” results (more useful knowledge graphs) for real problems applications;
- (ii) Identification of crucial research challenges and key points that should drive research efforts in the near future.

Based on the aforementioned items, some more discussions were motivated and the group tried to formulate possible tangible ways to cope with the representation problems. The main goal was to identify concrete ways to achieve better results from possible future collaborations and follow-up actions that might start after this seminar.

The summary of the last discussions is:

1. We can put together efforts already being done on information extraction (from different languages and also images) and knowledge graph building. One concrete actions include coupling Chinese information extraction to NELL system.
2. We can define a common application domain in which we could apply the results of the collaboration proposed in item 1 (above). One concrete example would be using information extraction and knowledge bases to identify events in sports games matches (i.e. penalty in a soccer match) and associate that event with social media (i.e. Twitter) real time discussions.
3. The open questions (that should motivate research efforts) are:
 - (i) How to evaluate the obtained results?
 - (ii) How to have a robust representation for different domains?
 - (iii) How to cope with temporal-spatial scope?
 - (iv) How to put together deeper ontologies (CyC) with shallow representations such as Knowledge graphs?

4.2 Working Group III: Visual Information and Knowledge Graphs

Dubravko Culibrk (University of Trento, IT)

License  Creative Commons BY 3.0 Unported license
© Dubravko Culibrk

Joint work of Witbrock, Michael;Grobelnik, Marko;Hodson, James;Pacifco, Stefano;Novak, Blaz

The working group discussions started with the examination whether the Knowledge Graphs (KGs) are in fact necessary and useful to aid the vision tasks? Eventually a consensus was reached that the existing Knowledge Bases (KBs), i.e. KGs do not have enough coverage of to help disambiguate and aid the vision tasks. Therefore the discussion from that point on

focused on how to achieve the required extension of the KBs and the density of KGs required to make them useful in scenarios of relevance to vision. To achieve this practically we thought we could limit ourselves to some specific tasks (such as understanding tabletops or meals) and extend existing KBs automatically. We would like to infer the world model for this domain, which could then be extended to other domains. We would like to model entities, relations, scripts. After the discussions with all the participants we extended our scenario to understanding everything that happens in a kitchen. The final discussions revolved around how we could go about setting up a challenge that would help build a dense model of the world of kitchen. We would need a lot of video data, which is publicly available. To help the participants we would like to have the automatic speech recognition transcripts for the video data in the dataset. The tasks would need to be defined to favour the teams that are able to effectively use the multimodal data and require the inference of the model. The final conclusion of this workgroup is that we would like to organise a challenge along the lines of has been discussed and pursue options to get EU funding to create and run the challenge over the next few years.

4.3 Working Group IV: Representation Learning

Aditya Mogadala (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license
© Aditya Mogadala

Joint work of Carreras, Xavier; Smeaton, Alan; Sebe, Nicu; Chong-Wah, Ngo; Thalhammer Andreas; Rettinger Achim

The working group discussions started with possible scenarios where multimedia can be leveraged with textual information and vice versa. Initially, an idea of using Google⁶ image search was pitched in to disambiguate textual queries of multimedia search. An idea of building on the fly image classifier by crawling first 1000 images from Google to disambiguate retrieved results. This paved the path in a new direction to understand the images which lack objects and are very abstract. For example, how can a computer understand images that means “emptiness”. This drove the discussions to explore external information provided in the form of text or structured knowledge. Possibilities of using structured knowledge was inspected to identify relationships between objects detected in media content. Representation learning can be used in this scenario as:

- Good approach to predict missing modality.
- Application driven.
- Don’t have to bother about features – deep learning

Other ideas that are brainstormed use both multimedia and textual information for aligning religious pictures present in the museums along with ancient texts. Most of the picture galleries present in museums either lack descriptions or difficult to interpret their inherent depth. There are many ancient texts written that would have described something similar in the pictures. Aligning picture galleries to ancient texts manually can be tedious and cumbersome. Leveraging multimedia processing approaches with natural language understanding techniques can automate this process to an extent. Few other ideas that was discussed are:

1. Identifying a dominant person in a video where two people are debating on a topic.
2. Identifying a person and his role in the news domain.

⁶ <https://google.com>

3. Predicting price from the product catalog.

4. Event identification about disasters

Further action was to come up with a benchmark dataset to solve any of these tasks and conduct challenge in future.

5 Open Problems

There are several open issues which needs to addressed. Few of them are listed below.

- Identifying if generative models work for cross-lingual and cross-media linking.
- What kind of approach that needs to be employed, if we do not have enough multilingual and multimedia training data.

Participants

- Xavier Carreras
Xerox Research Centre Europe –
Grenoble, FR
- Dubravko Culibrk
University of Trento, IT
- John Davies
BT Research – Ipswich, GB
- Tiansi Dong
Universität Bonn, DE
- Anastassia Fedyk
Harvard University, US
- Blaz Fortuna
Ghent University, BE
- Marko Grobelnik
Jozef Stefan Institute –
Ljubljana, SI
- Alexander G. Hauptmann
Carnegie Mellon University, US
- James Hodson
Bloomberg – New York, US
- Estevam R. Hruschka
University of São Carlos, BR
- Bea Knecht
Zattoo – Zürich, CH
- Juanzi Li
Tsinghua University – Beijing,
CN
- Dunja Mladenic
Jozef Stefan Institute –
Ljubljana, SI
- Aditya Mogadala
KIT – Karlsruher Institut für
Technologie, DE
- Chong-Wah Ngo
City University – Hong Kong,
HK
- Blaz Novak
Jozef Stefan Institute –
Ljubljana, SI
- Stefano Pacifico
Bloomberg – New York, US
- Achim Rettinger
KIT – Karlsruher Institut für
Technologie, DE
- Evan Sandhaus
The New York Times, US
- Nicu Sebe
University of Trento, IT
- Alan Smeaton
Dublin City University, IE
- Rudi Studer
KIT – Karlsruher Institut für
Technologie, DE
- Jie Tang
Tsinghua University – Beijing,
CN
- Andreas Thalhammer
KIT – Karlsruher Institut für
Technologie, DE
- Eduardo Torres Schumann
VICO – Leinfelden-Echterdingen,
DE
- Volker Tresp
Siemens – München, DE
- Christina Unger
Universität Bielefeld, DE
- Michael Witbrock
Cycorp – Austin, US
- Lexing Xie
Australian National University –
Canberra, AU
- Lei Zhang
KIT – Karlsruher Institut für
Technologie, DE



Theory of Evolutionary Algorithms

Edited by

Benjamin Doerr¹, Nikolaus Hansen², Christian Igel³, and
Lothar Thiele⁴

1 École Polytechnique – Palaiseau, FR, doerr@lix.polytechnique.fr

2 INRIA Saclay – Île-de-France – Orsay, FR, hansen@lri.fr

3 University of Copenhagen, DK, igel@diku.dk

4 ETH Zürich, CH, thiele@ethz.ch

Abstract

This report documents the talks and discussions at the Dagstuhl Seminar 15211 “Theory of Evolutionary Algorithms”. This seminar, now in its 8th edition, is the main meeting point of the highly active theory of randomized search heuristics subcommunities in Australia, Asia, North America, and Europe. Topics intensively discussed include rigorous runtime analysis and computational complexity theory for randomised search heuristics, information geometry of randomised search, and synergies between the theory of evolutionary algorithms and theories of natural evolution.

Seminar May 17–22, 2015- <http://www.dagstuhl.de/15211>

1998 ACM Subject Classification G.1.6 Optimization, F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases evolutionary algorithms, black-box optimization, randomized search heuristics, algorithms, artificial intelligence

Digital Object Identifier 10.4230/DagRep.5.5.57

Edited in cooperation with Carola Doerr

1 Executive Summary

Benjamin Doerr

Nikolaus Hansen

Christian Igel

Lothar Thiele

License © Creative Commons BY 3.0 Unported license
© Benjamin Doerr, Nikolaus Hansen, Christian Igel, and Lothar Thiele

Evolutionary algorithms (EAs) are randomized search and optimization methods applicable to problems that may be non-continuous, multi-modal, noisy, multi-objective or dynamic. They have successfully been applied to a wide range of real-world applications and have demonstrated impressive performance in benchmarks for derivative-free optimization. The seminar was devoted to the theory underlying evolutionary algorithms and related methods, in order to gain a better understanding of their properties and to develop new powerful methods in a principled way. The highly international, interdisciplinary seminar brought together leading experts and young researchers in the field. The 45 participants came from 13 different countries, spread over 4 continents. Many additional researchers had expressed their interest to also attend the seminar, but could unfortunately not be considered.



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

Theory of Evolutionary Algorithms, *Dagstuhl Reports*, Vol. 5, Issue 5, pp. 57–91

Editors: Benjamin Doerr, Nikolaus Hansen, Christian Igel, and Lothar Thiele



Dagstuhl Reports

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

Topics

The following report covers all important streams of research in the theory of evolutionary algorithms with a focus on three topics of particular current interest:

Runtime and complexity. Rigorous runtime and analysis and computational complexity theory have become the most important tools in the theory of discrete evolutionary algorithms. The Dagstuhl seminar series “Theory of Evolutionary Algorithms” has sparked this development. The drastic increase in new results, new methods, and young researchers entering this field, but also the major unsolved problems naturally lead to keeping this a focus topic.

Information geometry. Using concepts from information geometry in evolutionary algorithms is one of the most promising new theoretical direction in evolutionary computing. The seminar provided a unique opportunity to discuss perspectives and limitations of this approach.

Natural evolution. Evolutionary computing is rooted in theories of natural evolution, and many early approaches to understand basic properties of evolutionary algorithms were inspired by biological evolution theory. Still, today these two research fields are almost completely separated. We invited experts from evolution biology to help better understanding the relations between both fields. We are particularly happy that we succeeded in bringing together researchers from evolution biology and computer science in a way that was stimulating and productive.

Organization

The seminar had three types of organized presentation and discussion formats to stimulate the free discussions among the participants. There were 20–30 minutes talks on current topics followed by discussions. These included a talk on potential industrial collaborations. In addition, we had a few longer talks, which combined recent work with an overview over the state-of-the-art in a certain domain: Thomas Jansen spoke on “Understanding Randomised Search Heuristics”, Nick Barton on “Limits to Adaptation”, Yann Olivier introduced “Information-geometric Optimization”, and Timo Kötzing presented a talk on “Stochastic Fitness Functions and Drift”. Furthermore, we continued with having “breakout sessions” for longer, parallel group discussions on timely, specialized topics. These were introduced in the last seminar on “Theory of Evolutionary Algorithms”. This time, these sessions were even more productive than previously, both because the organizers and the participants were more used to this format of interaction. The talks and breakout sessions are summarized in Section 4 of this report.

We would like to thank the Dagstuhl team and the attendees for making seminar 15211 a great success and a pleasure to organize.

Benjamin Doerr (École Polytechnique – Palaiseau, FR)

Nikolaus Hansen (INRIA Saclay – Île-de-France – Orsay, FR)

Christian Igel (University of Copenhagen, DK)

Lothar Thiele (ETH Zürich, CH)

2 Table of Contents

Executive Summary

Benjamin Doerr, Nikolaus Hansen, Christian Igel, and Lothar Thiele 57

Overview of Talks

Spectral Landscape Theory: Some Color from Infinite Population Analysis <i>Lee Altenberg</i>	61
Limits to Adaptation <i>Nick Barton</i>	62
Fitness Information and Optimal Control of Mutation Rate <i>Roman V. Belavkin</i>	62
Information Geometric Optimization in \mathbb{R}^N : Open Problems and Misconceptions <i>Hans-Georg Beyer</i>	63
Selection of Auxiliary Objectives with Reinforcement Learning: Overview of Theoretical Results <i>Arina Buzdalova</i>	63
The Unrestricted Black-Box Complexity of Jump Functions <i>Maxim Buzdalov</i>	64
Understanding Simple Asynchronous Evolutionary Algorithms <i>Kenneth A. De Jong</i>	65
Analyzing Self-Adjusting Parameter Choices in Discrete Search Spaces <i>Carola Doerr</i>	65
On the Proportion of Fit Individuals in the Population of A Mutation-Based Genetic Algorithm <i>Anton V. Ereemeev</i>	66
Lyapunov Stability Analysis of a Derandomized Self-Adaptive (1, 2)-ES <i>Carlos M. Fonseca</i>	66
Probabilistic Analysis of the (1+1)-Evolutionary Algorithm <i>Hsien-Kuei Hwang</i>	67
Understanding Randomised Search Heuristics <i>Thomas Jansen</i>	67
Open Problems in Industry <i>Daniel Johannsen</i>	68
Stochastic Fitness Functions and Drift <i>Timo Kötzing</i>	68
Does the Natural Gradient Save Us? <i>Oswin Krause</i>	68
Level-based Analysis of Genetic Algorithms and Other Search Processes <i>Per Kristian Lehre</i>	69
Information Geometry of the Gaussian Distribution in View of Stochastic Optimization: Inverse Covariance Parameterization <i>Luigi Malago</i>	69

On the Runtime of Randomized Local Search and Simple Evolutionary Algorithms for Dynamic Makespan Scheduling <i>Frank Neumann</i>	70
Information-Geometric Optimization <i>Yann Ollivier</i>	70
Neutrality in Fitness Landscapes <i>Peter F. Stadler</i>	71
First Steps Towards a Runtime Comparison of Natural and Artificial Evolution <i>Dirk Sudholt</i>	71
The Potential of the Swarm’s Potential: Convergence, Stopping, Runtime, Stagnation <i>Rolf Wanka</i>	72
Population Size vs. Mutation Strength for the $(1+\lambda)$ EA on OneMax <i>Carsten Witt</i>	73
Working Groups	
Constrained Blackbox Optimization Benchmarking	73
Bridging the Gap Between Experiments and Theory Using Feature-Based Run-Time Analysis	78
Bringing together Evolutionary Computation and Population Genetics	79
Theory in Multimodal Optimisation	80
Theory of Genetic Programming	81
Continuous Optimization: Fitness vs. Ranks	84
Neutrality	84
Issues with Optimization for Machine Learning Using Variational Inference	85
Effects of Initialisation Process for Random Search Heuristics	85
Theory to Practice in Evolutionary Computation	86
Seminar Schedule	88
Participants	91

3 Overview of Talks

3.1 Spectral Landscape Theory: Some Color from Infinite Population Analysis

Lee Altenberg (*Konrad Lorenz Institute for Evolution & Cognition, AT*)

License  Creative Commons BY 3.0 Unported license
© Lee Altenberg

To obtain analytical results on runtimes of evolutionary algorithms, tractability has required the use of very simple landscapes. Is there any hope of obtaining results for arbitrary landscapes? There may perhaps be some relevance from models of evolutionary dynamics with infinite populations where powerful spectral methods have been able to provide a variety of results for arbitrary landscapes. Karlin (1982) proved two theorems that systems which combine growth and state transitions exhibit smaller aggregate growth with increased rates of transition. These theorems have been extended to infinite dimensional spaces, and multiple independent transition events (Altenberg, 2011, 2012). Their application has been used to show that dispersal should reduce standing genetic variation, that self-adaptation in stationary populations should favor lowered mutation and crossover rates (Feldman’s “Reduction Principle”), and that the fitness of a quasispecies decreases with increasing mutation rates up to $m=1/2$. In evolutionary computation, different representations and operators result in transmission matrices with different spectral properties. To illustrate, I show eight different ways to connect 64 genotypes by mutation, ranging from a simple path, to symbol sequences of length 6, 4, 3, and 2, to the complete graph, and show how the spectral gap of the associated mutation matrix increases for this series. Spectral landscape theory (Weinberger 1991, Stadler 1992) has produced a variety of results based on the relationships of 1) fitnesses to the 2) mutational eigenvalues and 3) mutational eigenvectors. Here I present new results showing that the asymptotic rate of growth of a quasispecies is an increasing function of the eigenvalues of its mutation matrix. In addition, a lower bound is found which increases with the alignment between fitnesses and the eigenvectors with largest eigenvalues. A principal open question most relevant to evolutionary computation is how the spectral gap of the quasispecies, which determines its rate of convergence from an initial state, depends on these three properties: fitnesses, mutational eigenvalues, and mutational eigenvectors.

References

- 1 Altenberg L (2011) An evolutionary reduction principle for mutation rates at multiple loci. *Bulletin of Mathematical Biology* 73:1227–1270.
- 2 Altenberg L (2012) Resolvent positive linear operators exhibit the reduction phenomenon. *Proceedings of the National Academy of Sciences U.S.A.* 109:3705–3710.
- 3 Feldman MW (1972) Selection for linkage modification: I. Random mating populations. *Theoretical Population Biology* 3:324–346.
- 4 Karlin S (1982) in *Evolutionary Biology*, eds Hecht MK, Wallace B, Prance GT (Plenum Publishing Corporation, New York) Vol. 14, pp 61–204.
- 5 Stadler PF (1992) Correlation in landscapes of combinatorial optimization problems. *Europhysics Letters* 20:479–482.
- 6 Weinberger ED (1991) Fourier and Taylor series on fitness landscapes. *Biological Cybernetics* 65:321–330.

3.2 Limits to Adaptation

Nick Barton (*IST Austria – Klosterneuburg, AT*)

License  Creative Commons BY 3.0 Unported license
© Nick Barton

What limits the effectiveness of selection? We have a good theoretical understanding of simple forms of selection, which is used both to optimise animal breeding, and to explain the prevalence of sex and recombination. This theory carries over directly to evolutionary computation. Complex “fitness landscapes” are less well understood, but nevertheless, there are some general results that constrain the rate at which selection can accumulate information.

3.3 Fitness Information and Optimal Control of Mutation Rate

Roman V. Belavkin (*Middlesex University, GB*)

License  Creative Commons BY 3.0 Unported license
© Roman V. Belavkin

Main reference R. V. Belavkin, “Dynamics of Information and Optimal Control of Mutation in Evolutionary Systems,” in A. Sorokin, R. Murphey, M. T. Thai, P. M. Pardalos (Eds.), “Dynamics of Information Systems: Mathematical Foundations,” Springer Proceedings in Mathematics & Statistics, Vol. 20, pp. 3–21, Springer, 2012.

URL http://dx.doi.org/10.1007/978-1-4614-3906-6_1

We consider genetic algorithms (GAs) as Markov chains, and then pose the problem of optimal control of their parameters. We give several formulations of this problem depending on additional constraints, such as time horizon and information constraints. In particular, we study the problem of optimal control of the mutation rate parameter. We show that solutions to the optimal control problems are control functions that can be derived from transition probabilities between level sets of the fitness function. Using our combinatorial result about the intersection of spheres in a Hamming space, we derive closed-form expressions for transition probabilities in the idealised case, when fitness is monotonic with respect to Hamming distance to an optimum [1]. The optimal mutation rate control functions are presented for several problems. We also discuss how these solutions can be applied to fitness functions that are only weakly monotonic. We also discuss the recent discovery of the mutation rate control in bacteria [2].

References

- 1 Belavkin, R. V. (2012). *Dynamics of Information and Optimal Control of Mutation in Evolutionary Systems*. In Sorokin, A.; Murphey, R.; Thai, M.T.; Pardalos, P.M. (Eds.) *Dynamics of Information Systems: Mathematical Foundations*. Springer Proceedings in Mathematics & Statistics, Vol. 20, pp 3–21. Springer.
- 2 Krasovec, R., Belavkin, R. V., Aston, J., Channon, A., Aston, E., Rash, B., Kadirvel, M., Forbes, S., Knight, C. (2014). *Mutation-rate-plasticity in rifampicin resistance depends on Escherichia coli cell-cell interactions*. *Nature Communications*, Vol. 5, No. 3742.

3.4 Information Geometric Optimization in \mathbb{R}^N : Open Problems and Misconceptions

Hans-Georg Beyer (Fachhochschule Vorarlberg, AT)

License © Creative Commons BY 3.0 Unported license
© Hans-Georg Beyer

Main reference H.-G. Beyer, “Convergence Analysis of Evolutionary Algorithms That are Based on the Paradigm of Information Geometry, *Evolutionary Computation*,” 22(4):679–709, 2014.

URL http://dx.doi.org/10.1162/EVCO_a_00132

After a short introduction into the philosophy and theory of Information Geometric Optimization (IGO) in real-valued search spaces, the dynamical theory of IGO on quadratic model functions will be reviewed. Recent results regarding the dynamics of the IGO flow [1] will be extended to provide convergence results for finite time steps in the case of expected value maximization. It will be shown that even in that case, the natural gradient definition leads to a disappointingly slow sublinear convergence order obeying an $1/t$ law. In order to get a faster approach (i.e., linear convergence order) to the optimizer one needs utility functionals that localize the IGO flow in time, such as local standardization of the sampling induced fitness values or truncation selection (i.e., local quantile based selection). Regarding the case of truncation selection, a special example will be considered where the fitness normality assumption [1] does not hold (therefore, the normality assumption is not the reason for the dynamical behaviors observed). The numerical solution of the resulting IGO ODE will be compared with the corresponding Evolution Strategy (ES). It will be shown that the classical mutation scale-invariance often observed in ES does not hold for a correctly working IGO algorithm unless one is willing to sacrifice the natural gradient definition and uses different time-step sizes for the mean and the covariance matrix update, respectively, thus departing from the IGO flow. The presentation will especially focus on open problems and misconceptions and is intended to initiate a vivid discussions regarding the concept of the natural gradient, information geometric optimization and its relation to Evolution Strategies.

References

- 1 H.-G. Beyer. Convergence Analysis of Evolutionary Algorithms That are Based on the Paradigm of Information Geometry. *Evolutionary Computation*, 22(4):679–709, 2014. DOI 10.1162/EVCO_a_00132.

3.5 Selection of Auxiliary Objectives with Reinforcement Learning: Overview of Theoretical Results

Arina Buzdalova (ITMO University – St. Petersburg, RU)

License © Creative Commons BY 3.0 Unported license
© Arina Buzdalova

Efficiency of evolutionary algorithms may be enhanced by multi-objectivization [1, 2]. One of the corresponding approaches is to dynamically select auxiliary objectives from some predefined set during one run of evolutionary algorithm (EA) [3]. Dynamic selection of objectives is needed when it is efficient to optimize different objectives at different stages of optimization. It was proposed to use reinforcement learning (RL) for dynamic selection of auxiliary objectives. The corresponding approach was called EA+RL [4]. This approach was shown to be efficient during experimental studies. However, there is need for theoretical understanding of the underlying mechanisms that enable EA+RL to work efficiently. In this

overview, recent theoretical results are considered [4, 5, 6]. Efficiency of EA+RL is evaluated on some easy problems using runtime analysis. Some insights on proper design choices when implementing EA+RL are described as well. Based on the considered results, we propose directions for future improvement of dynamic selection of auxiliary objectives in evolutionary algorithms.

References

- 1 J. D. Knowles, R. A. Watson, D. W. Corne. *Reducing Local Optima in Single-Objective Problems by Multi-objectivization*. In Proc. of the 1st Int'l Conf. on Evolutionary Multi-Criterion Optimization EMO'01, Springer, 2001, pp. 269–283.
- 2 C. Segura, C. A. C. Coello, G. Miranda, C. Léon. *Using multi-objective evolutionary algorithms for single-objective optimization*. 4OR, Vol. 3, No. 11, 2013, pp. 201–228.
- 3 Mikkel T. Jensen. *Helper-Objectives: Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimisation: Evolutionary Computation Combinatorial Optimization*. Journal of Mathematical Modelling and Algorithms, Vol. 3, No 4, 2004, p. 323–347.
- 4 M. Buzdalov, A. Buzdalova, A. Shalyto. *A First Step towards the Runtime Analysis of Evolutionary Algorithm Adjusted with Reinforcement Learning*. In Proc. of the Int'l Conf. on Machine Learning and Applications, Vol. 1, IEEE CS, 2013, pp. 203–208.
- 5 M. Buzdalov, A. Buzdalova. *OneMax Helps Optimizing XdivK: Theoretical Runtime Analysis for RLS and EA+RL*. In Proc. of Genetic and Evolutionary Computation Conf. Companion, ACM, 2014, pp. 201–202.
- 6 D. Antipov, M. Buzdalov, B. Doerr. *Runtime Analysis of (1+1) Evolutionary Algorithm Controlled with Q-learning using Greedy Exploration Strategy on OneMax+ZeroMax Problem*. Evolutionary Computation in Combinatorial Optimization, LNCS, 2015, Vol. 9026, pp. 160-1-72.

3.6 The Unrestricted Black-Box Complexity of Jump Functions

Maxim Buzdalov (ITMO University – St. Petersburg, RU)

License © Creative Commons BY 3.0 Unported license

© Maxim Buzdalov

Joint work of Buzdalov, Maxim; Doerr, Benjamin; Kever, Mikhail

Main reference M. Buzdalov, M. Kever, B. Doerr, “Upper and Lower Bounds on Unrestricted Black-Box Complexity of $\text{Jump}_{n,\ell}$,” in Proc. of the 15th Europ. Conf. on Evolutionary Computation in Combinatorial Optimization (EvoCOP'15), LNCS, Vol. 9026, pp. 209–221, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-16468-7_18

We analyse the unrestricted black-box complexity of the JUMP function class. For upper bounds, we present three algorithms for small, medium and extreme jump sizes. We prove a matrix lower bound theorem which is capable of giving better lower bounds than the classic information theory approach. Using this theorem, we prove lower bounds for JUMP separately for odd and even values of the bit string length. For several cases, notably for extreme jump sizes, lower and upper bounds coincide apart from lower order terms. Here is the complete list of proven upper bounds on $\text{JUMP}_{n,\ell}$:

- for $\ell < n/2 - \sqrt{n} \log_2 n$: $\frac{2n(1+o(1))}{\log_2 n}$, where $o(1)$ is measured when $n \rightarrow \infty$;
- for $n/2 - \sqrt{n} \log_2 n \leq \ell < \lfloor \frac{n}{2} \rfloor - 1$: $\frac{n(1+o(1))}{\log_2(n-2\ell)}$, where $o(1)$ is measured when $n - 2\ell \rightarrow \infty$;
- for $\ell = \lfloor \frac{n}{2} \rfloor - 1$: $n + \Theta(\sqrt{n})$.

The proven lower bounds are:

- for even n : $\left\lfloor \log_{\frac{n-2\ell+2}{2}} \left(1 + 2^{n-1} \frac{(n-2\ell)^2}{n-2\ell-1} \right) \right\rfloor - \frac{2}{n-2\ell} \geq \frac{n}{\log_2 \frac{n-2\ell+2}{2}} - 1$;
- for odd n : $\left\lfloor \log_{\frac{n-2\ell+1}{2}} \left(1 + 2^{n-2} (n-2\ell-1) \right) \right\rfloor - \frac{2}{n-2\ell-1} \geq \frac{n-1}{\log_2 \frac{n-2\ell+1}{2}} - 1$.

3.7 Understanding Simple Asynchronous Evolutionary Algorithms

Kenneth A. De Jong (George Mason University – Fairfax, US)

License © Creative Commons BY 3.0 Unported license
© Kenneth A. De Jong

Joint work of De Jong, Kenneth A.; Scott, Eric

Main reference E. Scott, K. A. De Jong, “Understanding Simple Asynchronous Evolutionary Algorithms”, in Proc. of the 2015 ACM Conf. on Foundations of Genetic Algorithms XIII (FOGA’15), pp. 85–98, ACM, 2015.

URL <http://dx.doi.org/10.1145/2725494.2725509>

When the cost of fitness evaluations is very high, parallel EAs are used to reduce the “wall clock” time required for an evolutionary run. In the case that there is high variance in the cost of these long running fitness evaluations, there can be a significant amount of multi-processor idle time when using synchronous parallel EAs. A simple asynchronous version is introduced that reduces idle time to zero, and analyzed it in several respects: 1) Does it have a built-in bias for individuals with faster running evaluation times? 2) For a fixed budget, how much improvement in wall clock time is obtained? 3) Is there also a speedup in finding better solutions faster?

3.8 Analyzing Self-Adjusting Parameter Choices in Discrete Search Spaces

Carola Doerr (CNRS and UPMC – Paris, FR)

License © Creative Commons BY 3.0 Unported license
© Carola Doerr

Joint work of Doerr Benjamin; Doerr, Carola

Main reference B. Doerr, C. Doerr, “Optimal Parameter Choices Through Self-Adjustment: Applying the 1/5-th Rule in Discrete Settings,” in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation (GECCO’15), pp. 1335–1342, ACM 2015; pre-print available as arXiv:1504.03212v1 [cs.NE], 2015.

URL <http://dx.doi.org/10.1145/2739480.2754684>

URL <http://arxiv.org/abs/1504.03212v1>

While evolutionary algorithms are known to be very successful for a broad range of applications, the algorithm designer is often left with many algorithmic choices, for example, the size of the population, the mutation rates, and the crossover rates of the algorithm. These parameters are known to have a crucial influence on the optimization time, and thus need to be chosen carefully, a task that often requires substantial efforts. Moreover, the optimal parameters can change during the optimization process. It is therefore of great interest to design mechanisms that dynamically choose best-possible parameters. An example for such an update mechanism is the one-fifth success rule for step-size adaption in evolutionary strategies. While in continuous domains this principle is well understood also from a mathematical point of view, no comparable theory is available for problems in discrete domains. In this work we show that the one-fifth success rule can be effective also in discrete settings. We regard the $(1 + (\lambda, \lambda))$ GA proposed in [1]. We prove that if its population size is chosen according to the one-fifth success rule then the expected optimization time on ONEMAX is linear. This is better than what *any* static population size λ can achieve and is asymptotically optimal also among all adaptive parameter choices.

References

- 1 B. Doerr, C. Doerr, and F. Ebel. *From black-box complexity to designing new genetic algorithms*. Theoretical Computer Science, 567:87–104, 2015.

3.9 On the Proportion of Fit Individuals in the Population of A Mutation-Based Genetic Algorithm

Anton V. Eremeev (Sobolev Institute of Mathematics – Omsk, RU)

License © Creative Commons BY 3.0 Unported license
© Anton V. Eremeev

Main reference A. V. Eremeev, “Modelling and Analysis of Genetic Algorithm with Tournament Selection,” in Proc. of the 4th Europ. Conf. on Artificial Evolution (AE’99), LNCS, Vol. 1829, pp. 84–95, Springer, 2000.

URL http://dx.doi.org/10.1007/10721187_6

In this talk, a fitness-level model of non-elitist mutation-only genetic algorithm (GA) with tournament selection is considered. The model provides upper and lower bounds for the expected proportion of the individuals with fitness above given thresholds. In the case of GA with bit-flip mutation and OneMax fitness function, the lower bounds are tight when population size equals one, while the upper bounds are asymptotically tight when population size tends to infinity. The lower bounds on expected proportions of sufficiently fit individuals may be obtained from the probability distribution of an appropriate generalized random walk. This approach yields polynomial upper bounds on the run-time of the Iterated version of the GA on 2-SAT problem and on a family of symmetric set cover problems proposed by E. Balas. The research is supported by Russian Foundation for Basic Research grant 15-01-00785.

3.10 Lyapunov Stability Analysis of a Derandomized Self-Adaptive (1, 2)-ES

Carlos M. Fonseca (University of Coimbra, PT)

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

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

The convergence of a simple derandomized self-adaptive (1, 2)-ES is investigated on the class of strictly unimodal functions of one variable that are symmetric about the optimum. Using the theoretical framework proposed by Semenov and Terkel [1], not only can a stability region for the self-adaptation parameters be analytically determined, but also upper bounds on the rate of convergence can be established, allowing appropriate values for the self-adaptation parameters to be obtained numerically. Simulation results are in agreement with the theoretical conclusions.

Acknowledgment. The authors acknowledge Brazilian agencies CAPES, CNPq and FAPEMIG, the Portuguese agency FCT, and the iCIS project (CENTRO-07-ST24-FEDER-002003), which is co-financed by QREN in the scope of the Mais Centro Program and European Union’s FEDER, for financial support.

References

- 1 Mikhail A. Semenov and Dmitri A. Terkel. *Analysis of convergence of an evolutionary algorithm with self-adaptation using a stochastic Lyapunov function*. Evolutionary Computation 11(4): 363–379, MIT Press, 2003.

3.11 Probabilistic Analysis of the (1+1)-Evolutionary Algorithm

Hsien-Kuei Hwang (Academica Sinica – Taipei, TW)

License © Creative Commons BY 3.0 Unported license
© Hsien-Kuei Hwang

Joint work of Hwang, Hsien-Kuei; Panholzer, Alois; Rolin, Nicolas; Tsai, Tsung-Hsi; Chen, Wei-Mei

Main reference H.-K. Hwang, A. Panholzer, N. Rolin, T.-H. Tsai, W.-M. Chen, “Probabilistic analysis of the (1+1)-evolutionary algorithm,” arXiv:1409.4955v1 [math.PR], 2014.

URL <http://arxiv.org/abs/1409.4955v1>

We give a detailed analysis of the cost used by the $(1 + 1)$ -evolutionary algorithm. The problem has been approached in the evolutionary algorithm literature under various views, formulation and degree of rigor. Our asymptotic approximations for the mean and the variance represent the strongest of their kind. The approach we develop is also applicable to characterize the limit laws and is based on asymptotic resolution of the underlying recurrence. While most approximations have their simple formal nature, we elaborate on the delicate error analysis required for rigorous justifications. Our main results state that on the fitness function OneMax, the cost used by $(1 + 1)$ -EA follows asymptotically a double-exponential distribution with mean and variance of order $n \log n$ and n^2 , respectively. On the other hand, if the fitness function is changed to LeadingOnes, then the cost satisfies a central limit theorem with quadratic mean and cubic variance. These results are not new but our methods of proof are not only rigorous but provide stronger asymptotic approximations.

3.12 Understanding Randomised Search Heuristics

Thomas Jansen (Aberystwyth University, GB)

License © Creative Commons BY 3.0 Unported license
© Thomas Jansen

Joint work of Corus, Dogan; He, Jun; Jansen, Thomas; Oliveto, Pietro S.; Sudholt, Dirk; Zarges, Christine

Main reference D. Corus, J. He, T. Jansen, P.S. Oliveto, D. Sudholt, C. Zarges, “On Easiest Functions for Somatic Contiguous Hypermutations And Standard Bit Mutations,” in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation Conference (GECCO’15), pp. 1399–1406, AMC, 2015..

URL <http://dx.doi.org/10.1145/2739480.2754799>

Different theoretical perspectives can help us understand different aspects of randomised search heuristics like evolutionary algorithms, artificial immune systems and others. One particularly successful theoretical perspective is run time analysis. Recently, He, Chen and Yao have described a way of identifying and constructing easiest and hardest functions (which imply smallest and longest expected run time, respectively) for elitist randomised search heuristics that have population size one. We use this idea to investigate the easiest function of the $(1 + 1)$ CHM, a variant of the $(1 + 1)$ EA where the standard bit mutations from the $(1 + 1)$ EA are replaced with somatic contiguous hypermutations from the B-Cell algorithm, a popular artificial immune system for optimisation. An easiest function for the $(1 + 1)$ CHM is an asymptotically hardest function for the $(1 + 1)$ EA. Motivated by their very different strengths and weaknesses we consider hybrid algorithms that combine both algorithms by applying one of the two mutation operators with fixed probability. We prove that these hybrids are efficient on easiest functions of their components. We also discuss that easiest functions for such hybrids are more difficult to find than one might believe. We conclude with a number of open questions that could yield additional insight into the performance of hybrid heuristics.

3.13 Open Problems in Industry

Daniel Johannsen (SAP Innovation Center – Potsdam, DE)

License  Creative Commons BY 3.0 Unported license
© Daniel Johannsen

SAP is one of the world’s leading companies in the market of enterprise resource planning systems. Two focus areas in SAP’s research strategy are Future of Knowledge Work and Personalized Medicine. Each of these areas is estimated to target a multi-trillion USD market by 2025. In the area of Future of Knowledge Work, the main challenge is to measure the business value of intangible assets of an organization such as employee skills and knowledge. Progress in this area is expected to be driven by self-adaptive and self-learning applications relying on algorithmic techniques such as evolutionary optimization and machine learning. Because of their versatility, the same techniques are also promising for the fundamentally different area of Personalized Medicine. We present six specific research challenges posed by different SAP business units. For each of these, SAP aims to establish long-term research collaborations with academic partners.

3.14 Stochastic Fitness Functions and Drift

Timo Kötzing (Hasso-Plattner-Institut – Potsdam, DE)

License  Creative Commons BY 3.0 Unported license
© Timo Kötzing
Joint work of Gießen, Christian; Friedrich, Tobias; Kötzing, Timo; Krejca, Martin; Sutton, Andrew

In this talk we will see the powerful method of drift analysis as a tool to analyze random processes. The two main drift theorems are the additive and the multiplicative drift theorem, which we will apply in exemplary cases both from continuous and discrete optimization. In the second part of the talk we will survey the current work on stochastic fitness functions. We will particularly discuss the disruptiveness of mutation operators and recent positive results on crossover operators.

3.15 Does the Natural Gradient Save Us?

Oswin Krause (University of Copenhagen, DK)

License  Creative Commons BY 3.0 Unported license
© Oswin Krause

Recently, optimizing the expectation of an objective function over some probability distribution by following the natural gradient of the distribution parameters became a popular approach to describe and implement Evolutionary Strategies. In this framework, samples are drawn from the distribution to estimate the gradient which is multiplied by the inverse of the Fisher Information Matrix to form the natural gradient. The natural gradient is usually argued to be superior to the “vanilla” gradient as it is independent of the parameterisation and because the direction of the natural gradient better reflects the actual change of the distribution. The freedom of parameterisation is especially regarded as beneficial as it fits to the notion of black-box optimization. In this talk, we show that the natural gradient is not necessarily unique and that we have a degree of freedom to change it, even without changing

the expectation or its gradient, challenging the idea of a black-box metric. This spotlight talk is designed to spark interesting discussions regarding the use of the natural gradient in black-box optimization.

3.16 Level-based Analysis of Genetic Algorithms and Other Search Processes

Per Kristian Lehre (University of Nottingham, GB)

License © Creative Commons BY 3.0 Unported license
© Per Kristian Lehre

Joint work of Corus, Dogan; Dang, Duc-Cuong; Eremeev, Anton

Main reference D. Corus, D.-C. Dang, A. V. Eremeev, P. K. Lehre, “Level-Based Analysis of Genetic Algorithms and Other Search Processes,” in Proc. of the 13th Int’l Conf. on Parallel Problem Solving from Nature (PPSN’14), LNCS, Vol. 8672, pp. 912–921, Springer, 2014.

URL http://link.springer.com/chapter/10.1007/978-3-319-10762-2_90

We describe a very general technique for proving upper bounds on the expected optimisation time of a broad class of non-elitist population-based optimisation processes, ranging from simple (μ, λ) EAs to more complex genetic algorithms (GAs), and estimation of distribution algorithms (EDAs) applied in uncertain environments.

References

- 1 Dogan Corus, Duc-Cuong Dang, Anton V. Eremeev, and Per Kristian Lehre. Level-based analysis of genetic algorithms and other search processes. In *Proceedings of the 13th International Conference on Parallel Problem Solving from Nature*, PPSN’14, pages 912–921. Springer International Publishing, 2014.
- 2 Duc-Cuong Dang and Per Kristian Lehre. Evolution under partial information. In *Proceedings of the 2014 Conference on Genetic and Evolutionary Computation*, GECCO ’14, pages 1359–1366, New York, NY, USA, 2014. ACM.
- 3 Duc-Cuong Dang and Per Kristian Lehre. Efficient optimisation of noisy fitness functions with population-based evolutionary algorithms. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, FOGA ’15, pages 62–68, New York, NY, USA, 2015. ACM.

3.17 Information Geometry of the Gaussian Distribution in View of Stochastic Optimization: Inverse Covariance Parameterization

Luigi Malago (Shinshu University, JP)

License © Creative Commons BY 3.0 Unported license
© Luigi Malago

Joint work of Malagò, Luigi; Pistone, Giovanni

Main reference L. Malagò, G. Pistone, “Information Geometry of Gaussian Distributions in View of Stochastic Optimization,” in Proc. of the 2015 ACM Conf. on Foundations of Genetic Algorithms XIII (FOGA’15), pp. 150–162, ACM, 2015; pre-print available as hal-01108986, 2015.

URL <http://dx.doi.org/10.1145/2725494.2725510>

URL <https://hal.inria.fr/hal-01108986>

We study the optimization of a continuous function by its stochastic relaxation, i.e., the optimization of the expected value of the function itself with respect to a density in a statistical model. In particular we focus on gradient descent techniques applied to the multivariate Gaussian distribution for the optimization of functions defined over continuous

domains. In the talk we present a parameterization for Gaussian distribution given by the mean and inverse covariance matrix based on the natural parameterization of the exponential family. We describe the advantages of this parameterization in the computation of natural gradient when a sub-model, based on conditional independence assumptions among variables, is employed in the relaxation. The use of restricted models instead of the full covariance plays an important role in the large scale setting, in particular when the interaction pattern between variables in the objective function is sparse. In the talk we will refer to standard results in the literature of the graphical models and in particular of Markov random fields. Based on a joint work with Giovanni Pistone, from Collegio Carlo Alberto.

3.18 On the Runtime of Randomized Local Search and Simple Evolutionary Algorithms for Dynamic Makespan Scheduling

Frank Neumann (University of Adelaide, AU)

License © Creative Commons BY 3.0 Unported license
© Frank Neumann

Joint work of Neumann, Frank; Witt, Carsten

Main reference F. Neumann, C. Witt, “On the Runtime of Randomized Local Search and Simple Evolutionary Algorithms for Dynamic Makespan Scheduling,” in Proc. of the 24th Int’l Joint Conf. on Artificial Intelligence (IJCAI’15), pp. 3742–3748, AAAI Press, 2015; pre-print available as arXiv:1504.06363v1 [cs.DS], 2015.

URL <http://arxiv.org/abs/1504.06363v1>

Evolutionary algorithms have been frequently used for dynamic optimization problems. With this paper, we contribute to the theoretical understanding of this research area. We present the first computational complexity analysis of evolutionary algorithms for a dynamic variant of a classical combinatorial optimization problem, namely makespan scheduling. We study the model of a strong adversary which is allowed to change one job at regular intervals. Furthermore, we investigate the setting of random changes. Our results show that randomized local search and a simple evolutionary algorithm are very effective in dynamically tracking changes made to the problem instance.

3.19 Information-Geometric Optimization

Yann Ollivier (University Paris Sud, FR)

License © Creative Commons BY 3.0 Unported license
© Yann Ollivier

Joint work of Ollivier, Yann; Akimoto, Youhei; Arnold, Ludovic; Auger, Anne; Hansen, Nikolaus

Main reference Y. Ollivier, L. Arnold, A. Auger, N. Hansen, “Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles,” arXiv:1106.3708v3 [math.OC], 2014.

URL <http://arxiv.org/abs/1106.3708v3>

In this talk I gave an introduction to information-geometric optimization (IGO). I started with how gradient descent is quite unsatisfactory due to great sensitivity to problem encoding (e.g., scaling of each coordinate). Moving to the space of probability distributions over the original search space can help, by evolving a probability distribution in an intrinsic, coordinate-independent way. I presented the IGO-ML (maximum likelihood) family of algorithms, which update a probability distribution based on sampling and increasing the log-likelihood of the best sampled points, similarly to the cross-entropy method but modified as to obtain coordinate-invariance. This applies both to discrete and continuous search

spaces, and recovers known algorithms from basic principles, such as PBIL (starting with Bernoulli distributions on the hypercube) and CMA-ES (starting with Gaussian distributions). Interestingly, IGO-ML has a guarantee of improvement for finite learning rates, as opposed to usual gradient descents for which such a guarantee requires infinitesimal steps. Moreover, from its information-geometric construction, IGO-ML has a guarantee of minimal change in diversity, which could help for multimodal optimization, as is confirmed by preliminary experiments.

3.20 Neutrality in Fitness Landscapes

Peter F. Stadler (Universität Leipzig, DE)

License © Creative Commons BY 3.0 Unported license
© Peter F. Stadler

Consider a fitness landscapes (X, f) . For simplicity we take the configuration space X as a graph, and f as real-valued function on the vertices. Concepts such as adaptive walks, gradients, etc. are well-defined in an obvious way whenever f is injective. Local failures of injectivity, i.e., the presence of adjacent points in X with the same values are usually referred to as neutrality. Degeneracies and neutrality come in at least two different flavors: (1) Degeneracy may be a consequence of coarse grained “increments” in the fitness function itself as in the case of NK models or spin glass models with small integer coefficients. (2) Degeneracy may also arise from intrinsic symmetries in the structure of the of the cost function, as in the case of short-range spin glasses. The consequence of neutrality, and possibly more general types of degeneracies is very poorly understood in general and both positive and negative effects of optimization have been reported: In the case of RNA landscapes, neutral networks seem to improve adaptability and search space exploration. Work by van Nimwegen & Crutchfield a decade ago, on the other hand described “entropic barriers”. Furthermore, ruggedness and neutrality are in principle independent properties and can be tuned in random landscape models independently of each other.

3.21 First Steps Towards a Runtime Comparison of Natural and Artificial Evolution

Dirk Sudholt (University of Sheffield, GB)

License © Creative Commons BY 3.0 Unported license
© Dirk Sudholt

Joint work of Paixão, Tiago; Pérez Heredia, Jorge; Sudholt, Dirk; Trubenová, Barbora

Main reference T. Paixão, J. Pérez Heredia, D. Sudholt, B. Trubenová, “First Steps Towards a Runtime Comparison of Natural and Artificial Evolution,” in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation (GECCO’15), ACM, pp. 1455–1462, 2015; pre-print available as arXiv:1504.06260v2 [cs.NE], 2015.

URL <http://dx.doi.org/10.1145/2739480.2754758>

URL <http://arxiv.org/abs/1504.06260v2>

Evolutionary algorithms (EAs) form a popular optimisation paradigm inspired by natural evolution. In recent years the field of evolutionary computation has developed a rigorous analytical theory to analyse their runtime on many illustrative problems. Here we apply this theory to a simple model of natural evolution. In the Strong Selection Weak Mutation (SSWM) evolutionary regime the time between occurrence of new mutations is much longer

than the time it takes for a new beneficial mutation to take over the population. In this situation, the population only contains copies of one genotype and evolution can be modelled as a (1+1)-type process where the probability of accepting a new genotype (improvements or worsenings) depends on the change in fitness. We present an initial runtime analysis of SSWM, quantifying its performance for various parameters and investigating differences to the (1+1) EA. We show that SSWM can have a moderate advantage over the (1+1) EA at crossing fitness valleys and study an example where SSWM outperforms the (1+1) EA by taking advantage of information on the fitness gradient.

3.22 The Potential of the Swarm’s Potential: Convergence, Stopping, Runtime, Stagnation

Rolf Wanka (*Universität Erlangen-Nürnberg, DE*)

License  Creative Commons BY 3.0 Unported license
© Rolf Wanka

Joint work of Bassimir, Bernd; Raß, Alexander; Schmitt, Manuel; Wanka, Rolf
Main reference B. I. Schmitt, “Particle Swarm Optimization Almost Surely Finds Local Optima,” Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, FAU University Press, 2015.
URL <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:29-opus4-61621>

After a short introduction into the continuous search heuristic “Particle Swarm Optimization” (PSO), the notion of the “potential” of the swarm is presented. The multifarious applications are shown: It can be shown that PSO in the 1-dimensional case converges towards a local optimum almost surely. As in the D -dimensional case ($D > 1$) the potential can be distributed unevenly among the dimensions, the movement equation is slightly changed to incorporate the potential. This can also be used to measure how often this rescue trick has to be used, in order to recognize that (presumably) the swarm is close to an optimum solution. Additionally, the potential can be used to define a measure of closeness to optimality such that, with the help of drift theory, a 1-dimensional runtime analysis can be performed. Finally, it can be explained in terms of the potential why sometimes the number of particles does not suffice to find a local optimum.

References

- 1 M. Schmitt. Convergence Analysis for Particle Swarm Optimization. Doctoral Thesis, University of Erlangen-Nuremberg, FAU University Press, 2015.[urn:nbn:de:bvb:29-opus4-61621]
- 2 M. Schmitt, R. Wanka. Particle Swarm Optimization Almost Surely Finds Local Optima. *Theoretical Computer Science*, 561A (2015) 57–72. [doi:10.1016/j.tcs.2014.05.017]
- 3 Rass, A.; Schmitt, M.; Wanka, R.: Explanation of Stagnation at Points that are not Local Optima in Particle Swarm Optimization by Potential Analysis. in: *Companion of Proc. 17th Genetic and Evolutionary Computation Conference(GECCO)*; 2015. [doi:10.1145/2739482.2764654]
- 4 B. Bassimir, M. Schmitt, R. Wanka. How Much Forcing is Necessary to Let the Results of Particle Swarms Converge? in: *Proc. Int. Conf. on Swarm Intelligence Based Optimization (ICSIBO)*, pp.98–105, 2014. [doi:10.1007/978-3-319-12970-9_11]

3.23 Population Size vs. Mutation Strength for the $(1+\lambda)$ EA on OneMax

Carsten Witt (Technical University of Denmark – Lyngby, DK)

License  Creative Commons BY 3.0 Unported license
© Carsten Witt

Joint work of Gießen, Christian; Witt, Carsten

Main reference C. Gießen, C. Witt, “Population Size vs. Mutation Strength for the $(1+\lambda)$ EA on OneMax,” in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation Conference (GECCO’15), pp. 1439–1446, ACM, 2015.

URL <http://dx.doi.org/10.1145/2739480.2754738>

The $(1+\lambda)$ EA with mutation probability c/n , where $c > 0$ is an arbitrary constant, is studied for the classical OneMax function. Its expected optimization time is analyzed exactly (up to lower order terms) as a function of c and λ . It turns out that $1/n$ is the only optimal mutation probability if $\lambda = o(\ln n \ln \ln n / \ln \ln \ln n)$, which is the cut-off point for linear speed-up. However, if λ is above this cut-off point then the standard mutation probability $1/n$ is no longer the only optimal choice. Instead, the expected number of generations is (up to lower order terms) independent of c , irrespectively of it being less than 1 or greater. The results are obtained by a careful study of order statistics of the binomial distribution and variable drift theorems for upper and lower bounds.

4 Working Groups

4.1 Constrained Blackbox Optimization Benchmarking

Organized by Dimo Brockhoff (INRIA and University of Lille 1, FR)

Participants: Youhei Akimoto, Dirk Arnold, Anne Auger, Hans-Georg Beyer, Dimo Brockhoff, Carlos Fonseca, Nikolaus Hansen.

4.1.1 Summary

The breakout session on constrained blackbox optimization benchmarking was one among several breakout sessions held during the Dagstuhl seminar 15211 “Theory of Evolutionary Algorithms”. Its main focus was to bring forward an extension of the well-established Blackbox Optimization Benchmarking exercise (BBOB) and its underlying Comparing Continuous Optimizers (Coco) platform towards constrained problems. The discussion covered in particular the concrete topics of designing constrained benchmark functions, performance assessment in the constrained case, and practical issues of benchmarking constrained blackbox optimization algorithms. This summary is expected to serve as a reference and the starting point to eventually provide an extension of the Coco platform towards constrained optimization to the research community.

4.1.2 Overview

The constrained blackbox optimization benchmarking breakout session took place in the evening (8:30pm–10:30pm) of the first day of the Dagstuhl seminar (May 18, 2015) and was attended by Youhei Akimoto, Dirk Arnold, Anne Auger, Hans-Georg Beyer, Dimo Brockhoff, Nikolaus Hansen, and Carlos Fonseca.

The breakout session started with a brief summary of known benchmarking exercises, the participants were aware of and used in their previous research (Sec. 4.1.3). Here, we will give a slightly extended overview of the state of the art in constrained optimization benchmarking. Later on, the discussion covered the topics of designing new benchmark functions, the interface between the test problem and the optimization algorithm (Sec. 4.1.4), and general aspects of performance assessment in the constrained case (Sec. 4.1.5). Throughout the discussion and this summary, we thereby assume the following constrained numerical optimization problem to be solved:

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{subject to } g_i(x) \geq 0 \text{ for all } 1 \leq i \leq m \end{aligned}$$

in which we call f the objective function (with continuous domain) and the g_i 's are called constraint functions.

4.1.3 State-of-the-Art Benchmarking

The state of the art in benchmarking numerical optimization algorithms on constrained problems centers around a few benchmark suites. Probably the most common benchmark suites in the context of blackbox optimization are related to the CUTE/CUTEr/CUTEst [1, 3, 4] framework and the competitions and special sessions organized at CEC 2006 [6] and CEC 2010 [7]. At CEC in 2008, 2010, 2015, additional competitions were held which covered bound constrained problems.¹

The latest CUTEst benchmark suite provides a general interface to more than 1000 test functions (specified in the Standard Input Format (SIF)). About 300 of them are constrained and scalable in the number of variables. Their constraints range from bound constraints over linear and quadratic constraints to general non-linear constraints. Both problems with inequality and equality constraints are present and the objective and constraint functions are separately callable. If available, derivatives are provided as well as known bounds on objectives and constraints. Besides implementations of the problems' objective and constraint functions and interfaces to several commonly used numerical optimization algorithms such as BOBYQA, Direct Search, or NEWUOA, the CUTE/CUTEr/CUTEst framework does not prescribe or provide facilities for performance assessment.

The latest CEC 2010 benchmark suite of constrained problems evolved from the earlier CEC 2006 testbed. It now contains 18 problems overall with a variety of constraint difficulties (bound constraints, inequality constraints, equality constraints). The objective functions are thereby often simple (e.g., the maximum over the variables; one third of the functions are separable) and the constraints can be relatively difficult (there are for example 18 equality constraints but only 21 inequality constraints present). Equality constraints are transformed into inequality constraints with a precision of $\epsilon = 10^{-4}$ by changing $g_i(x) = 0$ into $|g_i(x)| - \epsilon \leq 0$. The number of resulting inequality constraints per problem ranges from one to four. Regarding the performance assessment of optimization algorithms, the CEC 2010 competition prescribes to report statistics about the objective function values and the number of constraint violations over 25 algorithm runs at three predefined budgets in two dimensions (10-D and 30-D). Constraint function evaluations are thereby to be counted as any other objective function evaluation. Source code for the performance assessment or the creation of LaTeX tables is not provided.

¹ See http://www3.ntu.edu.sg/home/epnsugan/index_files/cec-benchmarking.htm for further details.

There are several other constrained benchmark functions, e.g., in the GLOBAL Library, the Hock/Schittkowski problems, Mittelmann's collection of AMPL problems, or in the COPS library, which, however, have not been discussed during the breakout session. For a more detailed list, we refer the interested reader to the web page of Arnold Neumeier at http://www.mat.univie.ac.at/~neum/glopt/test.html#test_constr.

All mentioned test suites and benchmarking exercises for constrained optimization have in common that they only provide implementations of the test functions and a rudimentary description of how to assess performance of algorithms (e.g., prescribed budgets of function evaluations at which the objective function values and constraint violations are supposed to be reported). Two exceptions are the (independently developed) frameworks PAVER2 [2] and optimizationBenchmarking.org [8], which are not specifically designed for constrained problems. Both frameworks allow to read in data from algorithm runs (in a standardized format) and to display the data in tables and data/performance profiles. The former framework, in addition, allows to detect inconsistencies in the data by running pre-defined tests. None of the available frameworks, however, allow for a “fully automated” benchmarking in which *all* the tedious and recurring steps of data acquisition, processing, and visualization are provided to the user.

As to *unconstrained* optimization, however, the state-of-the-art in automated algorithm benchmarking is more developed. A recent effort towards the goal of automated benchmarking is the Comparing Continuous Optimizers platform (Coco, see coco.gforge.inria.fr) and the corresponding Blackbox Optimization Benchmarking testbed and workshop series (BBOB, see the same link). The Coco platform currently offers two testbeds: one with 24 noiseless and one with 30 noisy test functions – all of them are unconstrained. The performance assessment is based on the well-established concept of target-based run lengths where the expected running time (ERT) to reach a certain target precision of the objective function is used as main performance criterion. The Coco platform is available in five different languages (C/C++, python, MATLAB/Octave, Java, and R). Algorithms, written in any of the provided languages, can be easily plugged to Coco by calling a generic objective function which both provides the function evaluation and collects the data for the performance assessment itself. A postprocessing script, written in python, then allows to display various plots and tables about the algorithm performance automatically up to the compilation of pre-prepared LaTeX templates.

Six BBOB workshops and special sessions have been organized around the Coco platform so far and the data sets of about 150 algorithm variants are available online. They can also all be used easily within the Coco postprocessing to compare the performance of algorithms and algorithm variants, see for example [5].

The participants of the breakout session had no doubt that a constrained BBOB testbed with the automated data gathering, performance assessment, and plotting functionality of the Coco platform will trigger a new interest in constrained blackbox optimization – especially in the communities of numerical blackbox optimizers and evolutionary computation. The resulting open access collection of easily comparable algorithm data will furthermore, as in the unconstrained BBOB case, allow eventually for recommendations on which algorithms to use for certain problem classes.

4.1.4 Benchmark Functions

In terms of the definition of a benchmark function suite, all participants quickly agreed on the following essential points:

- Functions shall be scalable with the search space dimension.
- Constraints in practice are often linear and/or of a blackbox type.
- There exist functions for which knowing the constraint function(s) $g(x)$ is crucial to find the optimum quickly (for example when the objective function's optimum lies in a very small basin of attraction but the constraint function(s) give(s) hints towards the optimum).
- The extension of the Coco platform should not distinguish between finding the feasible region and optimizing within it. An algorithm for constrained blackbox optimization should be able to do both simultaneously.

The discussion then focused on the general aspects of defining a concrete set of benchmark functions, especially with respect to which objective and constraint functions should be used when extending the BBOB testbed to the constrained case.

Given that already the sphere function with a single linear constraint seems difficult for many common evolutionary algorithms, the participants agreed on aiming to have three main function groups/testbeds in an initial constrained blackbox optimization exercise:

- linear orthogonal constrained problems (which includes the special case of bound constraints)
- problems with linear constraints which are not orthogonal
- problems with true blackbox constraints (some of the already established test problems of the BBOB testbed could be used as constraints)

Related to the previous topic was a discussion about the concrete interface between the constrained testbed and the optimization algorithm. In the cases of linear constraints, the breakout session's participants could imagine to allow the algorithm knowledge about the constraints' normal vectors and a support vector. Another option, working in all three above cases, is to expose all values $g_i(x)$ directly. The third option is to provide only a binary response about feasibility/infeasibility of a search point – even if this setting might be too restrictive for some search algorithms. In all cases, allowing the algorithm to make recommendations about its best estimate of the optimum seems to make sense like in the noisy case. A recommended point will thereby replace the last evaluated point for the performance assessment.

Though the discussion stayed focused around the mentioned topics and several concrete assumptions have been fixed throughout the breakout session, some important questions remain open: Probably the most important ones are related to the usage of the optimum in the performance assessment: in general, we do not always know the optimum analytically when using arbitrary constraint functions – even if we know the optimum of the (unconstrained) objective function. One open question is then, for example, how to use “the best function value known so far” as a reference in the performance assessment: it raises the question of how to approximate the best known value well and how to deal with potentially improving values over time. Note here that the constrained CEC benchmarks, mentioned earlier, report the best known value so far. However, the approach is not based on target-based run lengths as in the BBOB/Coco framework for which a good knowledge of the optimal function value seems more critical. Another open question is the relation between certain assumptions on the objective and constraint functions and existence guarantees of analytical descriptions of the optima.

As the first step towards a general constrained blackbox optimization benchmarking exercise, the participants finally thought that a first test suite should be defined, which does not include all existing BBOB functions but simpler ones for which the optima are

known also in the constrained case. This should be feasible under regularity and convexity assumptions.

4.1.5 Performance Assessment

The main difference between benchmarking unconstrained and constrained optimization algorithms is the way how the information about sampled infeasible solutions are used in the performance measures. Inherently, assessing performance in the constrained case can be defined as a multiobjective problem, with a trade off between the minimization of the objective function f and the minimization of the m constraint violations.

To not complicate the performance evaluation, the participants agreed that looking only at feasible search points in the performance measures seems to make sense as a first step when generalizing the Coco platform. This allows to use the same standard performance measures such as the expected running time (ERT) to hit a certain f -value target precision as in the unconstrained case. However, the number of used constraint evaluations should also be taken into account in addition to the number of function evaluations: in practice, an algorithm for constrained optimization might not even evaluate the objective function if the evaluation of the constraint functions already renders a search point infeasible. This is especially the case if both the objective and (some of) the constraint functions are of a blackbox type and expensive to compute.

Since it is important to know how the algorithm performs over time, in such a case, we need to know how the algorithm performance scales with the number of objective function evaluations *and* constraint function evaluations. We therefore suggest to have the axes in standard BBOB/Coco plots which are related to time (such as in the data profiles) as well as the definition of the expected running time to be defined according to either the number of objective function evaluations, the number of constraint function evaluations, or according to the sum of both. On the other hand, there might be cases, where both the computation of the objective function and the computation of the constraint functions are by themselves cheap but rely on some (joint) computations which are expensive (e.g., a numerical simulation). In this case, it seems reasonable to measure the search costs by the max of f -calls and g -calls, i.e., the number of the expensive simulations. All in all, we can expect that good algorithms for one case might not at all perform well in other cases. This is another good reason for pursuing the benchmarking efforts of the Coco framework to find out quantitative results. In order to do so, the benchmark suite might want to potentially offer the information about which objective function/constraint function combination is expensive to evaluate to the algorithms.

4.1.6 Conclusion

The breakout session on constrained blackbox optimization benchmarking covered the topics of designing new benchmark functions, the interface between the test problem and the optimization algorithm, and general fundamental aspects of performance assessment in the context of an extension of the well-established BBOB/Coco platform towards constrained optimization problems.

Beside making progress towards the concrete definition of a new constrained blackbox optimization benchmarking testbed based on the BBOB/Coco benchmarking exercise, the breakout session's discussion resulted in several new research questions – both of theoretical and practical nature. The breakout session has been very fruitful and has seen a lively and efficient discussion – as such, breakout sessions should therefore be kept in the program of future editions of the Dagstuhl seminar series.

Acknowledgments

Special thanks go to the organizers of the Dagstuhl seminar and the entire Dagstuhl team for providing such a stimulating and welcoming atmosphere.

References

- 1 I. Bongartz, A. R. Conn, N. Gould, and P. L. Toint. CUTE: Constrained and Unconstrained Testing Environment. *ACM Transactions on Mathematical Software (TOMS)*, 21(1):123–160, 1995.
- 2 M. R. Bussieck, S. P. Dirkse, and S. Vigerske. PAVER 2.0: An Open Source Environment for Automated Performance Analysis of Benchmarking Data. *Journal of Global Optimization*, 59(2–3):259–275, 2014.
- 3 N. I. M. Gould, D. Orban, and P. L. Toint. CUTer and SifDec: A Constrained and Unconstrained Testing Environment, revisited. *ACM Transactions on Mathematical Software*, 29(4):373–394, 2005.
- 4 N. I. M. Gould, D. Orban, and P. L. Toint. CUTEst: a Constrained and Unconstrained Testing Environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60(3):545–557, 2015.
- 5 N. Hansen, A. Auger, R. Ros, S. Finck, and P. Posík. Comparing Results of 31 Algorithms from the Black-box Optimization Benchmarking BBOB-2009. In J. Branke et al., editors, *GECCO workshop on Black-Box Optimization Benchmarking (BBOB’2010)*, pages 1689–1696. ACM, July 2010.
- 6 J. J. Liang, T. P. Runarsson, E. Mezura-Montes, M. Clerc, P. N. Suganthan, C. A. Coello Coello, and K. Deb. Problem Definitions and Evaluation Criteria for the CEC 2006 Special Session on Constrained Real-Parameter Optimization. Technical report, Nanyang Technological University, 2006.
- 7 R. Mallipeddi and P. N. Suganthan. Problem Definitions and Evaluation Criteria for the CEC 2010 Competition on Constrained Real-Parameter Optimization. Technical report, Nanyang Technological University, 2010.
- 8 T. Weise. optimizationbenchmarking.org, 2015. last accessed on June 17, 2015.

4.2 Bridging the Gap Between Experiments and Theory Using Feature-Based Run-Time Analysis

Organized by Frank Neumann (University of Adelaide, AU) and Heike Trautmann (Universität Münster, DE)

Feature-based analysis has been used to explore the performance of evolutionary algorithms based on features of given instances. Experimental studies show that problem instances may be classified as easy or difficult for a given EA based on instance features. Such classifications can be used for algorithm selection and/or parameter configuration. The goal of this breakout session was to discuss how mathematical analyses can contribute to this experimentally driven research area. A natural candidate is the runtime analysis of EAs taking into account problem features but any theoretical approach that supports these experimental investigations is highly welcome.

The session started with a brief introduction on feature-based analysis such that participants who were not familiar with this area could contribute to the discussion. The breakout session was targeted as a brainstorm session which should discuss approaches that may be applicable and there was the goal to identify some concrete problems that can become a topic of new research work.

The following challenges for a theoretical/mathematical analysis have been identified:

- for NP-hard problems: features will never capture all information about the problem
- which are the most important (ideally deterministic) features?
- features interact in practice

As a consequence, we think that an analysis should start using a single feature combined with a simple algorithm concept.

The break-out group identified the following topics for future research:

- start with classical polynomially solvable combinatorial problems
- examine structural instance features and influence on algorithm performance
- use empirical results in order to get hints to important features
- either runtime estimation directly or configuration of the algorithm to optimize runtime
- candidate problems: shortest path, spanning trees, matchings, graph colouring, k-Sat, finding subgraphs, knapsack, modified versions of OneMax

More specifically:

- Graph Colouring: Runtime (Phase transition) based on the density of the graph
- k-Sat: probability of failure based on ratio of variables and clauses
- knapsack: runtime based on profits and weights

Candidate algorithms for the investigations of these problems are simple algorithms such as the $(1+1)$ -EA and the $(1+\lambda)$ -EA. Based on the features, it would be interesting to optimize important parameters such as the mutation probability or offspring population size. It is highly recommended that the theoretical investigations utilize empirical studies for validating the alignment of theoretical findings with practical observations.

4.3 Bringing together Evolutionary Computation and Population Genetics

Organized by Tobias Friedrich (Hasso-Plattner-Institut, Potsdam, DE)

Participants: Lee Altenberg, Nick Barton, Arina Buzdalova, Carlos M. Fonseca, Tobias Friedrich, Kenneth A. De Jong, Joshua D. Knowles, William B. Langdon, Per Kristian Lehre, Jonathan L. Shapiro, Dirk Sudholt, Peter F. Stadler.

One of the three themes of this Dagstuhl meeting was “theory of natural evolution”. The meeting was joined not only by computer scientists working in evolutionary computation, but also by a number theoretical biologists and other related fields.

At the beginning of evolutionary computation, a lot of inspiration was drawn from biology. However, we discussed that for example the infinite population of Michael Vose is not considered useful for real GAs. It was argued that also in theoretical biology simplified models are essential. Some operators like crossover are very interesting for both communities. The differences often come from the different aims (explaining nature vs. optimization). There are still surprisingly many similarities. Another joint aspect are island models. Even evolutionary algorithms which appear to be rather different from nature, like genetic programming, have similar features: The common problem of bloat in genetic programming is also well known for genes in nature.

4.4 Theory in Multimodal Optimisation

Organized by Christine Zarges (University of Birmingham, GB)

Participants: Hans-Georg Beyer, Carola Doerr, Tobias Glasmachers, Thomas Jansen, Oswin Krause, Pietro S. Oliveto, Günter Rudolph, Christine Zarges.

4.4.1 Summary

The main aim of this breakout session was to discuss potential routes for theoretical research in multimodal optimisation and how theory could help to promote this important application area. The discussion centered around three different aspects: the goal of multimodal optimisation, benchmark functions for theory and the evaluation and comparison of different algorithms. Given the composition of the group, aspects relevant for both discrete and continuous domains were discussed.

4.4.2 What is the goal of Multimodal Optimisation?

There has been some discussion about the appropriateness of the term *multimodal optimisation* and what it incorporates. Multimodal optimisation is concerned with problems that have a number of different local and global optima. However, there are different optimisation goals one could be interested in. On one hand, one can take a global perspective and consider the goal of finding a single optimum of the given problem. For this case there are already a couple of theoretical results available, in particular in the context of pseudo-Boolean and combinatorial optimisation. On the other hand, one could be interested in finding several different optima, either in a simultaneous or sequential fashion. It was agreed that this latter perspective should be called *multi-local optimisation*. Different evaluation criteria in the context of multi-local optimisation are discussed in Section 4.4.4.

4.4.3 How does an appropriate benchmark look like?

In the context of discrete optimisation some analysis has been performed on the TWOMAX function, which can be defined as follows:

$$\begin{aligned} \text{TWOMAX} &: \{0, 1\}^n \rightarrow \mathbb{R} \text{ with} \\ \text{TWOMAX}(x) &= \max(|x|_0, |x|_1), \end{aligned}$$

where $|x|_0$ and $|x|_1$ denote the number of zeros and ones in x , respectively. This function has two global optima and different algorithms including diversity preserving mechanisms have been considered in its context. However, it has been agreed that a good benchmark function should have a much larger number of optima. After some discussion, we came up with a generalisation of the TWOMAX function, which we call k -MAX. The function has k different optima which can be selected randomly or purposefully. All optima have two parameters: slope ($a_i > 0$) and offset ($b_i > 0$). All other search points have a fitness value depending on the distance to the closest optimum such that each search point is in the basin of attraction of its closest optimum. k -MAX can formally be defined as follows:

$$\begin{aligned} k\text{-MAX} &: \{0, 1\}^n \rightarrow \mathbb{R} \text{ with} \\ k\text{-MAX}(x) &= n - \left(a_i \cdot H(x, x^{(i)}) + b_i \right) \quad \text{with } i = \operatorname{argmin}_{j \in \{1, \dots, k\}} H(x, x^{(j)}) \end{aligned}$$

where $H(x, x^{(i)})$ denotes the Hamming distance of search point x to optimum $x^{(i)}$. It is noted that this function exhibits a similar ‘constructive pattern’ (namely random locations of local optima and scaling) as Shekel’s foxholes in continuous optimisation.

4.4.4 How should different algorithms be compared?

Based on the k -MAX function, several algorithms were discussed. As usual algorithms can be compared under a runtime (*How long does your algorithm need to achieve a given goal?*) or fixed budget (*Given a fixed budget, how well can your algorithm perform?*) perspective. However, with respect to the evaluation of algorithms in a multimodal context there are other important aspects to consider:

- How many optima did the algorithm find?
- What quality do the optima have?
- What is the distribution of the optima?

A crucial insight is that there is not a single solution to the problem of algorithm evaluation since an appropriate metric depends on the concrete application at hand. The group agreed that evaluation criteria are an important question for future research. For some of these criteria and for several optimisation problems, the state-of-the-art technique in real-world optimisation challenges are (*random or deterministic*) restarts. The group agreed that it is an interesting unsolved question to classify problems for which population-based algorithms are superior to such restart strategies.

4.5 Theory of Genetic Programming

Organized by Pietro S. Oliveto (University of Sheffield, GB)

Participants: Lee Altenberg, Maxim Buzdalov, Arina Buzdalova, Benjamin Doerr, Thomas Jansen, Timo Kötzing, William B. Langdon, Alberto Moraglio, Frank Neumann, Pietro S. Oliveto, Dirk Sudholt, Carsten Witt, Christine Zarges.

4.5.1 Summary

The goal of Genetic Programming (GP) is to evolve computer programs i.e., executable functions with desired functionality. While numerous successful applications of GP have been reported, there is limited and fragmented theoretical understanding of the working principles of GP. Compared to the theoretical analysis of genetic algorithms, the study of GP poses several extra challenges, eg., tree representations of variable length. Some preliminary work on the theory of GP has been undertaken by various participants of this Dagstuhl seminar. The purpose of this breakout session was to discuss potential feasible and promising routes to build a theoretical foundation of GP, taking advantage of the progresses in the runtime analyses of evolutionary algorithms for function optimisation.

4.5.2 How to approach the theoretical analysis of GP?

The theoretical understanding of Evolutionary Algorithms (EAs) has grown considerably in the past 20 years with the development of mathematical techniques that have allowed runtime analyses of more and more sophisticated EAs for problems bearing increasing similarities with those tackled in real-world applications. The achievement of such results has been possible by gradually building upon techniques constructed through the analysis of extremely

simplified EAs, such as the notorious (1+1) EA, for artificially constructed test problems with particular characteristics. Popular examples of such problems are the Onemax and Trap functions that have characteristics that make the optimisation respectively easy and hard for EAs. Given the successful progress made over the years by the EA community, a promising direction to build a theoretical foundation of GP would be to follow a similar route to that taken by the field for the runtime analysis of EAs for function optimisation. The first steps in following such a strategy would be to identify the analogues for GP of the (1+1) EA and of useful test problems such as Onemax and Trap.

4.5.3 What should our GP algorithms evolve?

While the original motivation behind Genetic Programming (GP) was to introduce a tool that would allow the evolution of computer programs (i.e., executable functions with a desired functionality), GP techniques have also been used to address various kinds of other problems. Apart from evolving functions depending on input/output relationships, other perspectives have been to use GP to address problems that are typical to the field of machine learning or to evolve *designs* with a particular behaviour. Also, several different ways to represent a computer program have been proposed in the literature. Apart from the syntax trees used in the original GP, various linear representations have been proposed in subclasses of GP such as linear genetic programming or cartesian genetic programming. Since the use of syntax trees is the most wide spread, it was agreed that it is sensible to consider this kind of representation in the early foundational steps of GP theory.

Recent work concerning geometric semantic GP (GSGP), a non-standard form of GP, was also considered. Here the search operators are designed to induce a Onemax fitness landscape for any Boolean function, thus facilitating the obtainment of runtime analysis results [5]. It was remarked though that GSGP, even if efficient in the training phase, cannot possibly generalise well (at least on Boolean functions). Another research direction that was proposed was the design of refined search operators for GSGP that are provably efficient in the training phase and provably good at generalisation (in the PAC-learning sense) on some PAC-learnable sub-classes of all Boolean functions.

4.5.4 Test problems for GP

Since the goal of GP is to evolve functions with a desired functionality, an essential characteristic for a GP test problem is that the solution should be an executable function and qualities of candidate solutions should depend on how well they map inputs to outputs. Initial foundational steps should identify test problems that are particularly easy and particularly hard to evolve for a GP system. Such problems should have characteristics that will facilitate the progressive development of mathematical techniques to allow the analysis of more sophisticated algorithms for the same problems and of simple GP algorithms for more complicated problems. Several GP benchmarks were considered including the Max problem, even-n-parity, multiplexor problems [4], Royal trees [6, 7], tree-string problems [2] and K landscapes [8].

A promising domain is that of Boolean functions, especially because they are often used as benchmark problems in experimental GP research. Parity is well known to be a hard problem in the evolvability and GP communities, while other Boolean functions such as AND or ON (i.e., the Boolean function that always returns 1 as output) are known to be easy. A desirable characteristic for easy problems might be the lack of epistasis – that the change in fitness by flipping any bit is independent of the state of the bits at other loci, a

characteristic present in the Onemax function. No final conclusions were made concerning what the analogue of the Onemax test problem for GP is.

4.5.5 The Simple-GP Algorithm

Genetic Programming algorithms behave similarly to standard genetic algorithms (GAs) in the sense that they both have high crossover rates and very small mutation rates. A significant difference though is that GP trees vary in size and depth during the evolutionary process while GA genotypes remain of fixed size during the entire evolution. Hence, while the operators of a (1+1) GP may be simplified in a similar fashion to those of the (1+1) EA and random local search, they should allow trees to shrink and grow in size. The operators used in the few runtime analyses available in the literature [1, 3] (i.e., Insert, Substitute and Delete) seem suitable for the purpose. In these works symptoms of bloat have appeared even though no crossover was used.

4.5.6 Runtime Analysis

In order to determine how the computational resources depend on the size of the problem a notion of scalability has to be decided. Several ways could be used to represent the size of the problem. In some works the number of nodes in the optimal tree was used. The number of input variables of the benchmark function seems to be the most general and appropriate measure. Generalisation is an important issue to be considered in the analysis. The target function may not necessarily be the exact function (i.e., one of the optimal solutions) but one that approximates it nicely.

References

- 1 G. Durrett, F. Neumann, and U.-M O'Reilly. Computational complexity analysis of simple genetic programming on two problems modeling isolated program semantics. In *Proc. of the 11th Workshop on Foundations of Genetic Algorithms (FOGA '11)*, pages 69–80. ACM Press, 2011.
- 2 S. Gustafson, E. Burke, and N. Krasnogor. The tree-string problem: An artificial domain for structure and content search. In *Proceedings of EuroGP 2005*, pages 215–226. Springer, 2005.
- 3 T. Kötzing, A. M. Sutton, F. Neumann, and U.-M. O'Reilly. The max problem revisited: the importance of mutation in genetic programming. *Theoretical computer science*, 545:91–107, 2014.
- 4 W. B Langdon and R. Poli. *Foundations of Genetic Programming*. Springer, 2001.
- 5 A. Moraglio, A. Mambrini, and L. Manzoni. Runtime analysis of mutation-based geometric semantic genetic programming on boolean functions. In *Proceedings of the 12th workshop on foundations of genetic algorithms (FOGA '13)*, pages 119–131. ACM Press, 2013.
- 6 B. Punch, D. Zongker, and E. Goodman. The royal tree problem, a benchmark for single and multiple population genetic programming. In P. Angeline and K. Kinnear, editors, *Advances in Genetic Programming 2*, pages 299–316. MIT Press, 1996.
- 7 Walter Alden Tackett. *Recombination, selection, and the genetic construction of computer programs*. PhD thesis, University of Southern California, 1994.
- 8 L. Vanneschi, M. Castelli, and L. Manzoni. The k landscapes: a tunably difficult benchmark for genetic programming. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation (Gecco 2011)*, pages 1467–1474. ACM Press, 2011.

4.6 Continuous Optimization: Fitness vs. Ranks

Organized by Oswin Krause (University of Copenhagen, DK)

The current state of the art of continuous evolutionary optimization is to perform steps based on value comparisons and ranking of points, in contrast to of using the function values directly. This strategy, also known as “function value free” optimization, is very successful in practice as it gives invariance to monotone transformations of the objective function. This is in contrast to other randomized search techniques like stochastic gradient descent.

During the session we focused on the issue of noise handling. We concluded that proper ranking becomes hard once we are close to the optimum, especially when the noise is additive. In this case, ranking will always introduce a bias into the gradient estimate, which can (depending on the algorithm) slow down or prevent convergence to the optimum.

Therefore we discussed whether it is better to increase the population size or to reevaluate points in order to decrease the bias. Algorithmically, there is a clear trade-off between number of re-evaluations of an point and the number of steps taken, given a fixed budget of function evaluations, and it is in general unclear whether the bias induced by noise will prevent convergence or just slow it down (this depends on many aspects, e.g., how gradients are used, whether the noise is additive or multiplicative or something else, etc). In contrast, fitness-based algorithms (in contrast to rank-based, during the debate we borrowed the term “fitness proportional selection” to make the distinction explicit, although this changed the original meaning of the term) can easily be implemented to be unbiased and to converge (see standard results on (stochastic) gradient descent, SGD), while the trade-offs are not so important due to the fact that increasing the population size always decreases the noise on the gradient estimate, and noise can also average out over multiple steps (e.g., in standard SGD, due to the cooling schedule).

Still, removing ranks and working with plain fitness values in this case leads to a number of new problems as we loose the invariance against monotone transformations. Thus the algorithm needs to adapt to the curvature of the target function.

We discussed, whether there are useful transformations of the objective function other than ranking, however the easily computable ones are inferior and also need to be estimated from the noisy samples. It also has to be taken care that the transformation does not cause the vector field to contain circles, which can prevent convergence.

4.7 Neutrality

Organized by Peter F. Stadler (Universität Leipzig, DE)

Neutrality is clearly still a poorly understood topic. In biological landscapes, neutrality and redundancy does appear at sometimes extensive scales. It remains open under which conditions neutrality, or more precisely neutral networks, are beneficial to the performance of (EC-type) optimization algorithms. While examples are known where this is the case – using a non-uniform representation of phenotypes that favours good solutions, it is unclear whether uniform representations, in which each phenotype is represented by the same number of genotypes, can also have an advantage. Computational studies by Fonseca, for example, suggest that this is at least hard to achieve. An related question is whether neutrality can effectively be replaced by suitably expanded definitions of neighborhood. Conversely, there are combinatorial optimization problems, such as certain scheduling problems, which do not seem to have a essentially non-redundant or even non-neutral representations. So far, no

convincing strategies to construct efficient EC-type algorithms for the class of problems seems to be available. A point made repeatedly is the distinction between redundancy (degeneracy) of the fitness function and neutrality (i.e., local redundancy). This is also related to the issue of accessibility among the (phenotypic) representations. These can be effectively non-symmetric, for instance in the case that the redundancy of the encoding is very different for different phenotypes. To what extent this asymmetry is relevant to optimization algorithms remains open. The importance of neutrality in continuous settings remains unclear.

4.8 Issues with Optimization for Machine Learning Using Variational Inference

Organized by Yann Ollivier (CNRS and University Paris Sud, FR)

In this session we discussed specific technical problems that arise in some optimization problems from machine learning. In machine learning, typical optimization problems require finding the parameters of (probabilistic) models to best fit available data. However, there is a risk of overfitting, namely, finding a parameter value that is too precise given the available data. This is especially relevant in high dimension.

Avoiding the overfitting problem can be done by working with a modified objective, in which one is looking, not only for a single optimal parameter value, but for a whole region in parameter space where performance is good. In the “variational inference” approach, one works with probability distributions over the original parameter space, and the modified objective is the average of the original objective under this distribution, minus the entropy of the distribution. The latter term prevents collapsing to a single point. The whole modified objective has an interpretation as a compressed length of the data.

This setting is very similar to ones appearing in evolutionary strategies such as CMA-ES or NES. However, the specific form of the objective (an average of the original objective function) makes it impossible to directly use these strategies; especially, rank reweighting is impossible, which makes it technically much more difficult to update, e.g., a Gaussian distribution in a stable way.

Moreover, in high dimension, it is tempting to include a model selection aspect such as keeping or removing each individual component of the parameter, which puts a discrete optimization problem on top of the continuous parameter search. Once more, one must deal with an entropy term describing the information cost of selecting/deselecting components.

In the session we discussed possible approaches for these problems, but no fully satisfying solution was found during the time of the session, so that more work is clearly needed.

4.9 Effects of Initialisation Process for Random Search Heuristics

Organized by Jonathan E. Rowe (University of Birmingham, GB)

It is known that in some situations (e.g., $(1+1)$ EA on OneMax) that the exact method of initialisation does not have a significant effect on the run time (in this case, at most affecting low order terms). Empirically, however, it has been observed that for some problems the initialisation makes a big difference. One example is Set Cover, where it seems better to start with the empty set than with a random set. One reason for this is that a random set is likely to be feasible (at least on typical benchmark problems). The initial choice then

limits the area of the search space explored. Moreover, the fitness function gives no guidance as to which are the best elements to exclude – all choices improve fitness by one. On the other hand, the empty set is infeasible, and there will be selective pressure towards choosing sets which reduce the constraint violation. It therefore seems that the dependency on initial conditions arises through asymmetries in the information the fitness function provides for feasible and infeasible solutions. It would be an interesting research goal to analyse this effect theoretically. A number of other examples were discussed where a similar effect might be observed (e.g., TSP, Vertex Cover). One example where the analysis may be tractable is Minimum Spanning Tree, since much is already known about the run-time – although initialisation here is likely to only affect low order terms.

4.10 Theory to Practice in Evolutionary Computation

Organized by Joshua D. Knowles (University of Manchester, GB)

Participants: Dimo Brockhoff, Arina Buzdalova, Kenneth A. De Jong, Carlos M. Fonseca, Tobias Friedrich, Hsien-Kuei Hwang, Joshua D. Knowles, W. B. Langdon, Frank Neumann, Jonathan L. Shapiro.

Our question was: “How can we continue to bridge the gap between the practical application of EAs and the Theory of EAs?” In practical applications, algorithms are designed and tested to obtain maximum performance on complicated real-world optimization, learning and control problems; in theoretical work, the algorithms and problems are often simple in order to admit analysis and focus on core issues. Establishing a bridge from one end of this spectrum to the other is a central challenge of the Theory field (see e.g. [12]) and a core aim of the Theory of EAs series at Dagstuhl. In this breakout session we focused the discussion more particularly on identifying what kinds of practice work make a good target for theoretical analysis, and on the group’s personal experiences trying to bring practice and theory closer together.

We began by noting that in the wider Dagstuhl meeting a continuing bridging activity is apparent, and in diverse contexts. For example, we have seen rigorous theoretical work about hybrid algorithms consisting of more than one base algorithm² [11], use of auxiliary objective functions with adaptive switching between them [1, 3], asynchronous EAs for handling variability in the objective function latency [10], and several new analyses of sophisticated algorithms based on information geometry [5, 15]. Thus, while theoretical tools are still being sharpened in simpler settings, they are also, it seems, being used ‘in anger’ to obtain new results of more practical import.

Joshua Knowles next presented joint work on evolutionary algorithms for problems embedded in a resource-constrained context where evaluations of solutions are done by real (e.g., physical / chemical) experiments. In this general setting, certain parts of the feasible solution space are not available for some periods during the optimization run [2] because resources specific to the solution are required for its evaluation and may be depleted (e.g., in a problem to optimize combinatory drug therapies [18], certain drugs may be used up). Knowles suggested this sort of setting would be ripe for theoretical analysis because the optimization function itself can be simple (e.g., OneMax and TwoMax have been used in [2]), allowing a transfer of known results and techniques, yet, at the same time, empirical

² Also referred to as Algorithm Portfolios in some works.

results show that the effects of the resource constraints are interesting and can depend on a number of factors. Thus there is plenty in the area to challenge existing theoretical tools.

This raises the question to what extent do practical problems help theory? We did not reach a general answer to this but it was noted by Carlos Fonseca that sometimes practical problems do reinvigorate interest in a latent theoretical result in the literature. He gave the example that a known complexity bound for computing the hypervolume of the union of axis-aligned polytopes from computational geometry [14, 16] had been improved ([8, 19, 9, 7]) as a direct result of interest in evolutionary algorithms using hypervolume computations for assessing the population ‘fitness’ in multiobjective problems [13, 6, 4]. In general, however, it can be hard to get theoreticians interested in a practical problem and they may be unwilling to help even if they would be capable of doing so!

Kenneth A. De Jong made the point that it is crucial to find the right level of abstraction in describing problems (to make them accessible and interesting to theory) – very much an art rather than a science. Even this may not be enough to simulate immediate interest. Frank Neumann related recent experience in presenting a new benchmark, the Traveling Thief Problem (TTP) [17], which models an important aspect of many real-world problems that they are composed of two or more different combinatorial optimization subproblems (each of which may be relatively easy to solve in isolation). Even though TTP has this practical relevance, it seems difficult, at the present time, to make progress on it from a theoretical perspective and interest has not been piqued. Nevertheless, the existence of the challenge in a well-defined form, with much unimportant detail abstracted away, will hopefully provide impetus for theoretical attacks on it in time.

We concluded by confirming that the aim of bringing together theory and practice was essential to the continuing development of evolutionary methods, and one that should be pursued by practitioners as well as theorists. It is difficult to get the alignment to work – for the practical problem to be one that is ripe and interesting for theory – but we must not give up on this project. Indeed the only way to make this happen is to redouble our efforts to engage with and include practitioners (be it industrial or academic) in theory meetings in future.

References

- 1 Denis Antipov, Maxim Buzdalov, and Benjamin Doerr. Runtime Analysis of $(1 + 1)$ Evolutionary Algorithm Controlled with Q-learning Using Greedy Exploration Strategy on OneMax+ ZeroMax Problem. In *Evolutionary Computation in Combinatorial Optimization*, pages 160–172. Springer, 2015.
- 2 Richard Allmendinger and Joshua Knowles. On handling ephemeral resource constraints in evolutionary search. *Evolutionary Computation*, 21(3):497–531, 2013.
- 3 Arina Buzdalova and Maxim Buzdalov. Increasing efficiency of evolutionary algorithms by choosing between auxiliary fitness functions with reinforcement learning. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 1, pages 150–155. IEEE, 2012.
- 4 N. Beume. S-Metric Calculation by Considering Dominated Hypervolume as Klee’s Measure Problem. *Evolutionary Computation*, 17(4):477–492, 2009.
- 5 Hans-Georg Beyer. Information geometry optimization in \mathbb{R}^n : Open problems and misconceptions. In *Dagstuhl Reports (this issue)*. 2015.
- 6 N. Beume and G. Rudolph. Faster S-Metric Calculation by Considering Dominated Hypervolume as Klee’s Measure Problem. Technical Report CI-216/06, Sonderforschungsbereich 531 Computational Intelligence, Universität Dortmund, 2006. shorter version published at IASTED International Conference on Computational Intelligence (CI 2006).

- 7 K. Bringmann. Bringing order to special cases of klee’s measure problem. In *Mathematical Foundations of Computer Science (MFCS 2013)*, pages 207–218. Springer, 2012.
- 8 T. M. Chan. A (Slightly) Faster Algorithm for Klee’s Measure Problem. *Computational Geometry*, 43(3):243–250, 2010.
- 9 T. M. Chan. Klee’s Measure Problem Made Easy. In *Foundations of Computer Science (FOCS 2013)*, pages 410–419, 2013.
- 10 Kenneth A. De Jong. Understanding asynchronous evolutionary algorithms. In *Dagstuhl Reports (this issue)*. 2015.
- 11 Thomas Jansen. Understanding randomised search heuristics. In *Dagstuhl Reports (this issue)*. 2015.
- 12 Thomas Jansen and R. Paul Wiegand. Bridging the gap between theory and practice. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 61–71. Springer, 2004.
- 13 Joshua Knowles and David Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *Evolutionary Computation, IEEE Transactions on*, 7(2):100–116, 2003.
- 14 V. Klee. Can the Measure of $\bigcup_i^n \{a_i, b_i\}$ be Computed in Less Than $O(n \log n)$ Steps? *The American Mathematical Monthly*, 84(4):284–285, 1977.
- 15 Yann Ollivier. Laplace’s rule of succession in information geometry. *arXiv preprint arXiv:1503.04304*, 2015.
- 16 M. H. Overmars and C. K. Yap. New Upper Bounds in Klee’s Measure Problem. *SIAM Journal on Computing*, 20(6):1034–1045, 1991. conference version is oy1988a.
- 17 Sergey Polyakovskiy, Mohammad Reza Bonyadi, Markus Wagner, Zbigniew Michalewicz, and Frank Neumann. A comprehensive benchmark set and heuristics for the traveling thief problem. In *Proceedings of the 2014 conference on Genetic and evolutionary computation*, pages 477–484. ACM, 2014.
- 18 Ben G. Small, Barry W. McColl, Richard Allmendinger, Jürgen Pahle, Gloria López-Castejón, Nancy J. Rothwell, Joshua Knowles, Pedro Mendes, David Brough, and Douglas B. Kell. Efficient discovery of anti-inflammatory small-molecule combinations using evolutionary computing. *Nature Chemical Biology*, 7(12):902–908, 2011.
- 19 Hakan Yildiz and Subhash Suri. On Klee’s measure problem for grounded boxes. In *Symposium on Computational Geometry (SoGC 2012)*, pages 111–120. ACM, 2012.

5 Seminar Schedule

Monday

- | | |
|-------------|--|
| 9.00–10.00 | Introduction
Thomas Jansen: Understanding Randomised Search Heuristics (Part I) |
| 10.00–10.30 | Coffee Break |
| 10.30–12.00 | Thomas Jansen: Understanding Randomised Search Heuristics (Part II)
Hsien-Kuei Hwang: Probabilistic analysis of the (1+1)-evolutionary algorithm |
| 12.15–14.00 | Lunch Break |
| 14.00–15.30 | Kenneth A. De Jong: Understanding Asynchronous Evolutionary Algorithms
Proposals for Breakout Sessions |
| 15.30–16.00 | Coffee Break |
| 16.00–18.00 | Timo Kötzing: Introduction to Drift Analysis and Stochastic Optimization with Randomized Search Heuristics
Carsten Witt: Population Size vs. Mutation Strength for the (1+ λ) EA on OneMax |
| 20.30 | Breakout Session Dimo Brockhoff: Benchmarking & Constraints |

Tuesday

- 9.00–10.00 **Nick Barton:** Limits to adaptation
- 10.00–10.30 Coffee Break
- 10.30–12.00 **Peter F. Stadler:** Neutrality in Fitness Landscapes
Dirk Sudholt: First Steps Towards a Runtime Comparison of Natural and Artificial Evolution
- 12.15–14.00 Lunch Break
- 14.00–15.30 Breakout Session I
Frank Neumann and Heike Trautmann: Bridging the Gap Between Experiments and Theory Using Feature-Based Run-Time Analysis
Tobias Friedrich: Bringing together Evolutionary Computation and Population Genetics
Christine Zarges: Theory in Multimodal Optimisation
- 15.30–16.00 Coffee Break
- 16.00–18.00 **Frank Neumann:** On the Runtime of Randomized Local Search and Simple Evolutionary Algorithms for Dynamic Makespan Scheduling
Carola Doerr: Analyzing Self-Adaptive Parameter Choices in Discrete Search Spaces
Maxim Buzdalov: The Unrestricted Black-Box Complexity of Jump Functions

Wednesday

- 9.00–10.00 **Yann Olivier:** Information-geometric optimization with application to machine learning
- 10.00–10.30 Coffee Break
- 10.30–12.00 **Hans-Georg Beyer:** Information Geometry Optimization in \mathbb{R}^N
Open Problems and Misconceptions
Oswin Krause: Does the natural gradient save us?
- 12.15–13.30 Lunch Break
- 13.30–15.30 Hike
- 15.30–16.30 Coffee Break
- 16.30–18.00 Breakout Session II
Peter F. Stadler: Neutrality
Pietro S. Oliveto: Theory of Genetic Programming
Oswin Krause: Continuous Optimization: Fitness vs. Ranks

Thursday

- 9.00–10.00 **Arina Buzdalova:** Selection of Auxiliary Objectives with Reinforcement Learning: Overview of Theoretical Results
Carlos Fonseca: Lyapunov Stability Analysis of a Derandomized Self-Adaptive (1,2)-ES
- 10.00–10.30 Group Photo and Coffee Break
- 10.30–12.00 **Anton Eremeev:** On the proportion of fit individuals in the population of a mutation-based genetic algorithm
Per Kristian Lehre: Level-based Analysis of Genetic Algorithms and Other Search Processes
- 12.15–14.00 Lunch Break
- 14.00–15.30 Breakout Session III
Yann Olivier: Issues with Optimization for Machine Learning Using Variational Inference
Jonathan Rowe: Effects of Initialisation Process for Random Search Heuristics
Joshua Knowles: Theory to Practice in Evolutionary Computation
- 15.30–16.00 Coffee Break
- 16.00–18.00 **Daniel Johansen:** Open Problems in Industry
Summary and Discussion of Breakout Sessions

Friday

- 9.15–10.15 **Lee Altenberg:** Spectral Landscape Theory: Some Color from Infinite Population Analysis
Rolf Wanka: The Potential of the Swarm's Potential
- 10.15–10.45 Coffee Break
- 10.45–12.00 **Roman V. Belavkin:** Fitness Information and Optimal Control of Mutation Rate
Luigi Malagó: Information Geometry of the Gaussian Distribution in View of Stochastic Optimization: Inverse Covariance Parameterization
- 12.15 Lunch

Participants

- Youhei Akimoto
Shinshu University, JP
- Lee Altenberg
Konrad Lorenz Institute for
Evolution & Cognition, AT
- Dirk V. Arnold
Dalhousie University, CA
- Anne Auger
INRIA Saclay –
Île-de-France, FR
- Nick Barton
IST Austria –
Klosterneuburg, AT
- Roman V. Belavkin
Middlesex University, GB
- Hans-Georg Beyer
Fachhochschule Vorarlberg, AT
- Dimo Brockhoff
INRIA – University of Lille 1, FR
- Maxim Buzdalov
ITMO University –
St. Petersburg, RU
- Arina Buzdalova
ITMO University –
St. Petersburg, RU
- Kenneth A. De Jong
George Mason University –
Fairfax, US
- Benjamin Doerr
Ecole Polytechnique –
Palaiseau, FR
- Carola Doerr
UPMC – Paris, FR
- Anton V. Eremeev
Sobolev Inst. of Mathematics –
Omsk, RU
- Carlos M. Fonseca
University of Coimbra, PT
- Tobias Friedrich
Hasso-Plattner-Institut –
Potsdam, DE
- Tobias Glasmachers
Ruhr-Universität Bochum, DE
- Nikolaus Hansen
INRIA Saclay – Île-de-France –
Orsay, FR
- Hsien-Kuei Hwang
Academica Sinica – Taipei, TW
- Christian Igel
University of Copenhagen, DK
- Thomas Jansen
Aberystwyth University, GB
- Daniel Johannsen
SAP Innovation Center –
Potsdam, DE
- Joshua D. Knowles
University of Manchester, GB
- Timo Kötzing
Hasso-Plattner-Institut –
Potsdam, DE
- Oswin Krause
University of Copenhagen, DK
- Marvin Künnemann
MPI für Informatik –
Saarbrücken, DE
- William B. Langdon
University College London, GB
- Per Kristian Lehre
University of Nottingham, GB
- Luigi Malago
Shinshu University – Nagano &
INRIA Saclay
- Silja Meyer-Nieberg
Universität der Bundeswehr –
München, DE
- Alberto Moraglio
University of Exeter, GB
- Frank Neumann
University of Adelaide, AU
- Pietro S. Oliveto
University of Sheffield, GB
- Yann Ollivier
University Paris Sud, FR
- Jonathan E. Rowe
University of Birmingham, GB
- Günter Rudolph
TU Dortmund, DE
- Jonathan L. Shapiro
University of Manchester, GB
- Peter F. Stadler
Universität Leipzig, DE
- Dirk Sudholt
University of Sheffield, GB
- Lothar Thiele
ETH Zürich, CH
- Heike Trautmann
Universität Münster, DE
- Rolf Wanka
Univ. Erlangen-Nürnberg, DE
- Carsten Witt
Technical University of Denmark
– Lyngby, DK
- Xin Yao
University of Birmingham, GB
- Christine Zarges
University of Birmingham, GB



Multi-disciplinary Approaches to Reasoning with Imperfect Information and Knowledge – a Synthesis and a Roadmap of Challenges

Edited by

Igor Douven¹, Gabriele Kern-Isberner², Markus Knauff³, and Henri Prade⁴

- 1 Paris-Sorbonne University, FR, igor.douven@paris-sorbonne.fr
- 2 TU Dortmund, DE, gabriele.kern-isberner@cs.tu-dortmund.de
- 3 Universität Giessen, DE, markus.knauff@psychol.uni-giessen.de
- 4 Paul Sabatier University – Toulouse, FR, prade@irit.fr

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15221 “Multi-disciplinary Approaches to Reasoning with Imperfect Information and Knowledge – a Synthesis and a Roadmap of Challenges”.

This multi-disciplinary seminar brought together researchers from computer science, philosophy and psychology dealing with topics of rational reasoning, reasoning with imperfect information and rational decision making in real world problems. The different views from computational, logical and cognitive perspectives provided new insights on overlapping goals and complementary questions, for instance, when psychologists being interested in new formal models and computer scientists being interested whether their developed methods are materially adequate discussed logical and terminological backgrounds with philosophers. The combination of introductory talks, presentations and discussions of current work of the participants and discussion groups dealing with general questions lead to fruitful discussions where challenges for new paradigms of rational reasoning as well as visions and foci for interdisciplinary work were raised.

Seminar May 25–29, 2015 – <http://www.dagstuhl.de/15221>

1998 ACM Subject Classification F.3.0 Logics and Meanings of Programs (general), I.2.0 Artificial Intelligence (general), I.2.4 Knowledge Representation Formalisms and Methods

Keywords and phrases uncertain reasoning, commonsense reasoning, rational reasoning, conditionals, nonmonotonic logics, belief revision, similarity-based reasoning: case-based, interpolative, extrapolative, and analogical reasoning, reasoning on time, space, and causality, abductive reasoning, semantical frameworks for reasoning, decision making, decision support systems, intelligent agents, multiagent systems, knowledge engineering, ontologies, description logics, formal epistemology, formal social epistemology, philosophy of probability, philosophy of uncertainty modeling, cognitive aspects of rational and uncertain reasoning, cognitive theories of uncertain reasoning, empirical studies on rational and uncertain reasoning, belief revision

Digital Object Identifier 10.4230/DagRep.5.5.92

Edited in cooperation with Christian Eichhorn



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

Multi-disciplinary Approaches to Reasoning with Imperfect Information and Knowledge – a Synthesis and a Roadmap of Challenges, *Dagstuhl Reports*, Vol. 5, Issue 5, pp. 92–114

Editors: Igor Douven, Gabriele Kern-Isberner, Markus Knauff, and Henri Prade



DAGSTUHL
REPORTS Dagstuhl Reports

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

1 Executive Summary

Igor Douven

Gabriele Kern-Isberner

Markus Knauff

Henri Prade

License © Creative Commons BY 3.0 Unported license
© Igor Douven, Gabriele Kern-Isberner, Markus Knauff, and Henri Prade

This multi-disciplinary seminar with attendees from computer science, philosophy, and psychology addressed typical problems that smart and intelligent systems in real-world scenarios have to deal with both from formal and empirical points of view. Such systems have to face, in particular, the problem of reasoning with uncertain, imprecise, incomplete, or inconsistent (in short, imperfect) information which often renders more classical, i.e., strict or deductive methods obsolete or fallacious. Reasoning with imperfect information plays a central role in practical deliberation and rational decision making. Models of human context-dependent reasoning that synthesise logical, philosophical and psychological aspects would be helpful for designing better systems. In psychology, an increasing interest in new formal methods for rational human reasoning under uncertainty can be observed, and on the other hand, philosophers and computer scientists have shown an increased attention to the experimental methods of psychology recently. In particular for computer scientists and AI researchers, it is becoming more and more interesting to see whether the systems they have been developing are materially adequate. A synthesis of rational reasoning with imperfect information that takes into account research done in artificial intelligence, but also in psychology and philosophy is needed for providing a clearer view of where we are and what are the pending issues both from computational resp. logical and cognitive viewpoints. This will help making intelligent systems more effective, and more helpful for their human users.

This seminar brought together researchers interested in rational and uncertain reasoning from a very broad scientific scope to present and discuss problems and approaches from different disciplines, consolidate common grounds, and initiate new interdisciplinary collaborations. The seminar took profit from the fact that computer scientists, philosophers, and psychologists have started quite recently to work in a common methodological paradigm with overlapping goals, converging interests, and largely shared research tools. The attendees identified challenges for new paradigms of rational reasoning, and discussed visions and foci for more interdisciplinary work.

The first day, the seminar started with (invited) survey talks on central cross-field topics, where each topic was addressed by two researchers from different disciplines:

- *Nonmonotonic reasoning and change of knowledge and beliefs*
Marco Ragni (CS/Psy), Hans Rott (Phil)
- *Uncertain reasoning and decision theory*
Wolfgang Spohn (Phil), Henri Prade (CS)
- *Argumentation and reasoning under inconsistency*
Ofer Arieli (CS), Ulrike Hahn (Psy)
- *General forms of human reasoning (e.g., analogical reasoning, interpolation, and extrapolation, case-based reasoning)*
Vittorio Grotto (Psy), Steven Schockaert (CS)

The schedule for the next days included both sessions where attendees could present and discuss their work with the audience, and time slots for discussion groups. The topics of the

discussion groups were discussed in a plenary session, and four groups came out of that:

- *Topics of group 1:* Philosophers' and psychologists' view on human reasoning, and what computer scientists can contribute to that; axiomatic systems vs. psychological models – how do they fit?
- *Topics of group 2:* Empirical implications of formal reasoning systems and vice versa
- *Topics of group 3:* Combination/mixture of reasoning methods, qualitative vs. quantitative approaches; formal axiomatic systems are suitable for decision making(?)
- *Topics of group 4:* Promises and problems of probability theory; reliability, coherence, higher order probabilities

Groups 1 and 2 joined after the first session due to the closeness of the discussed topics. On Friday morning, the results of the working groups were presented, and a final, lively discussion in the plenary session closed the seminar.

	Tuesday May 26th	Wednesday May 27th	Thursday May 28th	Friday May 29th
8:45/9:00	8:45 Opening Survey Talk 1	Working groups	Presentations	Presentations Working groups
10:30–11:00	Coffee break			
	Survey Talk 2	Presentations	Presentations	Presentations Working Groups Wrapping up
12:15–14:00	Lunch			
	Survey talk 3	Working groups	Presentations	
15:30–16:00	Coffe Break			
	Survey talk 4 17:30 Setting up working groups	Short walk bike ride, ...	Working Groups	
18:00	Dinner	Barbecue	Dinner	

2 Table of Contents

Executive Summary

Igor Douven, Gabriele Kern-Isberner, Markus Knauff, and Henri Prade 93

Overview of Talks

Desirable Properties of Paraconsistent Logics <i>Ofer Arieli</i>	97
Towards a dual process cognitive model for argument evaluation <i>Florence Bannay-Dupin de St-Cyr</i>	97
Non-Classical and Cross-Domain Reasoning <i>Tarek R. Besold</i>	97
Coherent uncertain reasoning <i>Nicole Cruz de Echeverria Loebell</i>	98
Abstract Girotto Talk <i>Vittorio Girotto</i>	98
Argumentation and reasoning under inconsistency (in psychology) <i>Ulrike Hahn</i>	99
Bayesian Argumentation <i>Stephan Hartmann</i>	99
An epistemic extension of equilibrium logic and its relation to Gelfond's epistemic specifications <i>Andreas Herzig</i>	100
Open conditionals in the light of dynamic epistemic logics <i>Andreas Herzig</i>	100
Short Introduction to Computational Models of Argument <i>Anthony Hunter</i>	100
The dual-strategy model of deductive inferences <i>Henry Markovits</i>	101
A prioritized assertional-based revision for DL-Lite knowledge bases <i>Odile Papini</i>	101
Coherence under uncertainty: Philosophical and psychological applications <i>Niki Pfeifer</i>	102
Why not just (Bayesian) probabilities? <i>Henri Prade</i>	102
Nonmonotonic reasoning <i>Marco Ragni</i>	103
Four floors for the theory of belief change (and in particular, the case of imperfect discrimination) <i>Hans Rott</i>	103
Human plausible reasoning as a model for robust inference from imperfect knowledge <i>Steven Schockaert</i>	103

Uncertain Reasoning and Decision Theory <i>Wolfgang Spohn</i>	104
Knowledge, Uncertainty, and Ignorance <i>Sara L. Uckelman</i>	105
Possible Worlds Semantics for Conditionals: The Case of Chellas-Segeberg Semantics <i>Matthias Unterhuber</i>	105
Knowledge and gossip <i>Hans Van Ditmarsch</i>	105
Back to the Future – On the State of the Art in Default Reasoning <i>Emil Weydert</i>	106
Dilation and Delayed Decisions <i>Gregory Wheeler</i>	106
Representation and Bayesian Rational Predication <i>Momme von Sydow</i>	107
Working Groups	
Working Group “Combining and comparing qualitative and quantitative approaches to decision theory” <i>Ofer Arieli, Christoph Beierle, Tarik Besold, Florence Bannay-Dupin de Saint-Cyr, Steven Schockaert, Wolfgang Spohn, Sara L. Uckelman, and Emil Weydert</i>	107
Working Group “Probability & Inconsistency” <i>Nicole Cruz, Ulrike Hahn, Stephan Hartmann, Karolina Krzyżanowska, Momme von Sydow, Matthias Unterhuber, and Greg Wheeler</i>	109
Working Group “Empirical Implications of formal systems for reasoning and vice versa” <i>Igor Douven, Christian Eichhorn, Thomas Eiter, Castaneda Gazzo, Estefania Lupita, Vittorio Girotto, Andreas Herzig, Anthony Hunter, Gabriele Kern-Isberner, Markus Knauff, Henry Markovits, Odile Papini, Niki Pfeifer, Henri Prade, Marco Ragni, Hans Rott, Henrik Singmann, and Hans Van Ditmarsch</i>	112
Open Problems	113
Participants	114

3 Overview of Talks

3.1 Desirable Properties of Paraconsistent Logics

Ofer Arieli (Academic College of Tel Aviv, IL)

License © Creative Commons BY 3.0 Unported license
© Ofer Arieli

Joint work of Arieli, Ofer; Avron, Arnon; Zamansky, Anna

Paraconsistent logics are logics that tolerate inconsistent information in a non-trivial way. However, it is not always clear what should be the exact nature of such logics, and how to choose one for a specific application.

In this talk, we formulate a list of desirable properties that a ‘decent’ paraconsistent logic should have, and investigate the relations between them. This is exemplified in the context of 3-valued semantics, which is the simplest and the most popular framework for reasoning with contradictory data.

3.2 Towards a dual process cognitive model for argument evaluation

Florence Bannay-Dupin de St-Cyr (Paul Sabatier University – Toulouse, FR)

License © Creative Commons BY 3.0 Unported license
© Florence Bannay-Dupin de St-Cyr

Joint work of Bisquert, Pierre; Croitoru, Madalina; Bannay-Dupin de St-Cyr, Florence

In this study we are interested in the computational and formal analysis of the persuasive impact that an argument can produce on a human agent.

We propose a dual process cognitive computational model based on the highly influential work of Kahnemann and investigate its reasoning mechanisms in the context of argument evaluation. This formal model is a first attempt to take a greater account of human reasoning and is a first step to a better understanding of persuasion processes as well as human argumentative strategies, which is crucial in collective decision making domain.

3.3 Non-Classical and Cross-Domain Reasoning

Tarek R. Besold (Universität Osnabrück, DE)

License © Creative Commons BY 3.0 Unported license
© Tarek R. Besold

Analogy and analogical reasoning is one of the most studied representatives of a family of non-classical forms of reasoning working across different domains.

In the first part of the talk, I will shortly introduce general principles of computational analogy models (relying on a generalization-based approach to analogy) and will have a closer look at Heuristic-Driven Theory Projection (HDTP) as an example for a theoretical framework and implemented system. HDTP has been applied to model a diverse range of phenomena including “classical” analogical reasoning, but also inductive generalization and concept formation in mathematics, transfer learning, or essential part of concept blending processes.

The second part will deal with some short reflections on the application of complexity theory and tractability considerations to (theoretical and/or computational) cognitive models, using HDTP as a worked out example. I will advocate the need for cognitive models and systems to be plausible also with respect to the required computational resources, suggesting parameterized complexity theory and approximation theory as sources of inspiration for analysis.

3.4 Coherent uncertain reasoning

Nicole Cruz de Echeverria Loebell (University of London, GB)

License © Creative Commons BY 3.0 Unported license
© Nicole Cruz de Echeverria Loebell

Joint work of Cruz, Nicole; Baratgin, Jean; Oaksford, Mike; Over, David

The psychology of reasoning has traditionally used binary logic to assess whether people's beliefs are consistent and their inferences valid. But most of our beliefs, premises, and conclusions in both everyday life and science are uncertain, and this uncertainty cannot be expressed in binary logic. The probabilistic approach to deductive reasoning proposes a generalisation of consistency for categorical beliefs to coherence for degrees of belief. We examined the coherence of people's probability judgments for a range of one-premise inferences with conditionals, conjunctions and disjunctions.

People's responses were coherent at above chance levels for all inferences investigated, with two qualifications. First, people committed the conjunction fallacy, violating coherence, when the and-elimination inference ($p \wedge q$, therefore p) was presented using the materials typically leading to the fallacy.

Second, people's responses were coherent above chance levels assuming that the conditional was interpreted as satisfying the Equation $P(\text{if } p \text{ then } q) = P(q|p)$; but responses were incoherent above chance levels assuming that the conditional was interpreted as the material conditional of binary logic, with $P(\text{if } p \text{ then } q) = P(\text{not} - p \text{ or } q)$.

3.5 Abstract Girotto Talk

Vittorio Girotto (University of Venezia, IT)

License © Creative Commons BY 3.0 Unported license
© Vittorio Girotto

Main reference L. Fontanari, M. Gonzalez, G. Vallortigara, V. Girotto, "Probabilistic cognition in two indigenous Mayan groups," in Proc. of the National Academy of Sciences, 111(48):17075–17080, 2014.

URL <https://dx.doi.org/10.1073/pnas.1410583111>

Correct probabilistic evaluations are one of the hallmarks of rationality. A classical view is that the ability to make them depend on formal education.

Following this view, individuals living in traditional cultures are unable to reason about probabilities, and premodern individuals lacked even the basic notions of probability. Another view, which one can trace back to John Locke, is that a sense of chance emerges regardless of instruction and culture. This talk reviews recent studies showing that young children and adults with no formal education are able to solve a variety of probabilistic problems. The talk discusses the implications of this finding for the question of the relation between normative and common reasoning.

3.6 Argumentation and reasoning under inconsistency (in psychology)

Ulrike Hahn (University of London, GB)

License  Creative Commons BY 3.0 Unported license
© Ulrike Hahn

On the one hand, there is a widespread sense that the human cognitive system is riddled with inconsistent and conflicting beliefs. At the same time, however, human cognitive flexibility in light of a noisy, changing environment far surpasses and machine system to date. The talk surveys work in a variety of fields within psychology on how human beings respond to conflicting and/or inconsistent information.

3.7 Bayesian Argumentation

Stephan Hartmann (LMU München, DE)

License  Creative Commons BY 3.0 Unported license
© Stephan Hartmann

I will motivate and sketch a Bayesian theory of argumentation. According to this theory, an agent has prior beliefs about some propositions A, B, \dots . These beliefs are represented by a probability distribution P . The agent then learn the premisses of an argument from some information source. She may, for example, learn that A is the case and that A implies B . This amounts to the following constraints on the agent's new probability distribution P' : $P'(A) = 1$ and $P'(B|A) = 1$. The full new probability distribution is then determined by minimization of the Kullback-Leibler divergence between P' and P . One then obtains $P'(B) = 1$ as one would expect from modus ponens. In a similar way, one can examine the inference patterns modus tollens, affirming the consequent, and denying the antecedent. This approach can be generalized in many respects. The agent may, for example, not fully trust the source that A is true and only assign a very high new probability to A (in the case of modus ponens). Or she may have beliefs about a disabling condition D that inhibits B . In this case the agent learns (or so I argue) that $P'(B|A, \neg D) = 1$ where the variable D has to be properly integrated into a causal Bayes net that represents that conditional independences that hold between the various variables. Finally one may want to study alternatives to the Kullback-Leibler divergence and explore what follows from these measures. All this will, or so I hope, nicely connect to empirical studies.

3.8 An epistemic extension of equilibrium logic and its relation to Gelfond’s epistemic specifications

Andreas Herzig (Paul Sabatier University – Toulouse, FR)

License © Creative Commons BY 3.0 Unported license
© Andreas Herzig

Joint work of Fariñas del Cerro, Luis; Herzig, Andreas; Iraz Su, Ezgi
Main reference L. Fariñas del Cerro, A. Herzig, E. Iraz Su, “Epistemic Equilibrium Logic,” in Proc. of the 24th Int’l Joint Conf. on Artificial Intelligence (IJCAI’15), pp. 2964–2970, 2015; pre-print available from author’s webpage.

URL <http://ijcai.org/papers15/Abstracts/IJCAI15-419.html>

URL <http://www.irit.fr/~Andreas.Herzig/P/Ijcai15.html>

We add epistemic modal operators to the language of here-and-there logic and define epistemic here-and-there models. We then successively define epistemic equilibrium models and autoepistemic equilibrium models. The former are obtained from here-and-there models by the standard minimisation of truth of Pearce’s equilibrium logic; they provide an epistemic extension of that logic. The latter are obtained from the former by maximising the set of epistemic possibilities; they provide a new semantics for Gelfond’s epistemic specifications. For both definitions we characterise strong equivalence by means of logical equivalence in epistemic here-and-there logic.

3.9 Open conditionals in the light of dynamic epistemic logics

Andreas Herzig (Paul Sabatier University – Toulouse, FR)

License © Creative Commons BY 3.0 Unported license
© Andreas Herzig

We argue that Public Announcement Logic accounts an open epistemic conditional and show that it validates the principle of Stalnaker’s basic conditional logic, while it invalidates all the further principles of Lewis’s sphere-system-based logic of conditionals but the principle $A > false \rightarrow (A \wedge A') > false$.

3.10 Short Introduction to Computational Models of Argument

Anthony Hunter (University College London, GB)

License © Creative Commons BY 3.0 Unported license
© Anthony Hunter

Argumentation is an important cognitive process for dealing with incomplete, inconsistent, and uncertain information, and for dealing with conflicting view between agents. Computational models of argument aim to formalize aspects of argumentation for use in software. In this talk, we will consider models based on abstract argumentation, logical argumentation and dialogical argumentation.

3.11 The dual-strategy model of deductive inferences

Henry Markovits (University of Montreal, CA)

License © Creative Commons BY 3.0 Unported license
© Henry Markovits

Joint work of Markovits, Henry; Janie Brisson, de Chantal, Pier-Luc

The ability to make deductive inferences, that is, to understand that a single conclusion is a logical consequence of whatever preconditions are assumed is possibly the highest form of human cognition. A great deal of evidence has found that the inferences made even by educated adults are highly variable, but that this variability is not random. Instead people show clear patterns that reflect the specific content of the premises used in deductive problems. This variability underlies the development of probabilistic theories of inferential reasoning (Evans, Over & Handly, 2005; Oaksford & Chater, 2007). Although specific details differ, these theories suggest that people construct a statistical estimation of the probability that a putative conclusion is true, given what they know about the real world and the given premises, and that this statistical estimation is then used to produce a deductive conclusion. The other principle theory of deductive reasoning is mental model theory (Johnson-Laird, 2001; Johnson-Laird & Byrne, 2002), which proposes in contrast that people use a relatively conscious, working-memory intensive process to make inferences. This theory suggests that people construct representations of problem premises that consist of a series of models corresponding to combinations of antecedent and consequent terms that are semantically possibly true. The key aspect of this approach is the idea that if a reasoner can generate an explicit representation that includes a counterexample to a putative conclusion, this conclusion will be rejected. Recently, the Leuven group (Verschueren, Schaeken, & d'Ydewalle, 2005a; 2005) proposed a dual process theory of deductive reasoning which claims that people can use a combination of probabilistic and mental model forms of reasoning. We have been trying to confirm and extend this basic theory. We have been able to develop a method for evaluating the strategy used by reasoners (Markovits, Lortie Forgues, & Brunet, 2012). We have demonstrated that when people have a limited time to make inferences, they will preferentially use a probabilistic strategy, but will change to a mental model strategy when allowed more time (Markovits, Brunet, Thompson & Brisson, 2013). We have also shown that responses to deductive updating problems vary according to strategy use (Markovits, Brisson, de Chantal, in press). These results provide strong support for the dual strategy model.

3.12 A prioritized assertional-based revision for DL-Lite knowledge bases

Odile Papini (University of Marseille, FR)

License © Creative Commons BY 3.0 Unported license
© Odile Papini

We investigate Prioritized Removed Sets Revision (PRSR) for revising DL-Lite knowledge bases when a new sure piece of information, called the input, is added. The strategy of revision is based on inconsistency minimization and consists in determining smallest subsets of assertions (removed sets) that should be dropped from the current knowledge base in order to restore consistency and accept the input. We consider a DL-Lite knowledge base

where the ABox is stratified, and we consider different form of input: membership assertion, positive or negative axiom. To characterize an revision approach, we rephrase the Hansson's postulates for belief basis revision within DL-Lite settings and we give the logical properties of PRSR operators

3.13 Coherence under uncertainty: Philosophical and psychological applications

Niki Pfeifer (LMU München, DE)

License © Creative Commons BY 3.0 Unported license

© Niki Pfeifer

Joint work of Pfeifer, Niki; Sanfilippo, Giuseppe; Gilio, Angelo

Main reference A. Gilio, N. Pfeifer, G. Sanfilippo, "Transitive Reasoning with Imprecise Probabilities," in Proc. of the 13th Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'15), LNCS, Vol. 9161, pp. 95–105, Springer, 2015.

URL https://dx.doi.org/10.1007/978-3-319-20807-7_9

After sketching selected philosophically and psychologically interesting key features of coherence-based probability logic, we illustrate our approach by inferences about conditionals and quantified statements. Specifically, we discuss Modus ponens, Modus tollens, Cut, Contraposition, and selected paradoxes of the material conditional. Moreover, we present first steps towards a coherence-based probability semantics of categorical syllogisms. Finally, we discuss the importance of managing zero antecedent probabilities for reasoning under uncertainty (uncertain conditionals, probabilistic existential import assumptions, etc.).

3.14 Why not just (Bayesian) probabilities?

Henri Prade (Paul Sabatier University – Toulouse, FR)

License © Creative Commons BY 3.0 Unported license

© Henri Prade

Joint work of Prade, Henri; Dubois, Didier

This attempt at providing a brief overview of uncertainty modeling in artificial intelligence starts by recalling some limitations of precise probabilities with respect to the representation of epistemic uncertainty. The settings of possibility, belief function, and imprecise probability theories that rely on the use of two dual set functions, are shown to be appropriate for modeling (partial) ignorance. Then the importance of a proper view of conditioning, via conditional objects, is stressed, together with its relation to nonmonotonic reasoning, and its application to perceived causality. The distinction between qualitative possibility vs. quantitative possibility theory that relies on different definitions of conditioning is then recalled. Motivations for decision criteria beyond expected utility are also briefly indicated in presence of epistemic uncertainty for one-shot decisions. Lastly, two recent research trends are briefly mentioned:

1. The structure of the cube of opposition that applies to possibility and belief function theories (as well as to many other knowledge representation frameworks) emphasizes the existence of two other set functions of interest in these two settings
2. The existence of a qualitative counterpart to belief function theory based on imprecise possibilities.

3.15 Nonmonotonic reasoning

Marco Ragni (Universität Freiburg, DE)

License © Creative Commons BY 3.0 Unported license
© Marco Ragni

In the last decades psychological findings show that human reasoning strongly deviates from classical logical approaches. Nonmonotonic logics provide a better predictability of logical inferences. In this talk I will first introduce a variety of formal nonmonotonic reasoning approaches, from Reiter's Default Logic, System P, System C and focus especially on promising semantic and syntactic approaches. Accompanying questions about psychological demands for an adequate cognitive nonmonotonic theory are discussed.

3.16 Four floors for the theory of belief change (and in particular, the case of imperfect discrimination)

Hans Rott (Universität Regensburg, DE)

License © Creative Commons BY 3.0 Unported license
© Hans Rott

Main reference H. Rott, "Four Floors for the Theory of Theory Change: The Case of Imperfect Discrimination", in Proc. of the 13th Europ. Conf. on Logics in Artificial Intelligence (JELIA'14), LNCS, Vol. 8761, pp. 368–382, Springer, 2014.

URL http://dx.doi.org/10.1007/978-3-319-11558-0_26

The classical qualitative theory of belief change due to Alchourrón, Gärdenfors and Makinson has been widely known as being characterized by two packages of postulates. While the basic package consists of six postulates and is very weak, the full package that adds two further postulates is very strong. I revisit two well-known constructions of belief contraction, viz., contraction based on possible worlds and entrenchment-based contraction, and argue that four intermediate levels can be distinguished that play – or ought to play – important roles within qualitative belief revision theory. Levels 3 and 4 encode two ways of interpreting the idea of imperfect discrimination of the plausibilities of possible worlds or propositions.

3.17 Human plausible reasoning as a model for robust inference from imperfect knowledge

Steven Schockaert (Cardiff University, GB)

License © Creative Commons BY 3.0 Unported license
© Steven Schockaert

An increasing amount of structured knowledge is becoming available on the web (e.g. extracted from natural language). Unfortunately, classical logic is ill-suited for dealing with the uncertainty, vagueness, subjectivity and context-dependence that is prevalent in knowledge bases which have been derived from the web. In general, problems tend to arise whenever we need to reason about knowledge that has been encoded by humans (e.g. regulations, expert systems, ontologies). The challenge in reasoning about human knowledge is two-fold. First, such knowledge tends to capture statistical regularities (i.e. observations about the world) rather than tautologies. This is addressed, for example, in probabilistic extensions to classical logic (e.g. Markov logic), as well as in frameworks for non-monotonic reasoning. Second,

many of the concepts and properties about which we need to reason cannot be adequately defined using necessary and sufficient conditions. This has led to the use of geometric models of reasoning, inspired by cognitive theories of meaning (e.g. prototype theory), in particular various forms of similarity and analogy based reasoning.

Existing approaches for reasoning about human knowledge can broadly be categorised based on how they handle uncertainty (i.e. the first challenge) and concept representation (i.e. the second challenge), and based on whether they deal with these challenges in a qualitative or in a numerical way. Qualitative approaches have the advantage that their reasoning processes are more transparent, and accordingly, that intuitive explanations for inference results can easily be provided. Qualitative knowledge bases are also easier to learn, as there is no need for context-specific weights to be chosen. However, qualitative approaches are often too cautious in practice, which means that most existing applications rely on numerical, often heuristic approaches to reasoning. Numerical approaches, on the other hand, have to rely on weights, which may be difficult to learn in a principled way. This becomes especially problematic in approaches that handle both uncertainty and concept representation in a numeric way. While such approaches have already proven useful in applications, their knowledge bases rely on weights which are highly context-specific, and are difficult to maintain as a result.

The limited transferability of weighted knowledge, along with the difficulty to generate faithful and intuitive explanations for inference results from such knowledge, is likely to become increasingly problematic, as artificial intelligence methods are becoming increasingly central to human decision support. For example, doctors are unlikely to put much faith in computer-generated diagnoses, when they are based on imperfect methods, unless they can verify the reasoning process behind them. Similarly, regulators may insist on some degree of transparency when automated methods are used e.g. for approving mortgages, deciding insurance premiums, or assessing job candidates. An important challenge thus consists in developing methods that combine the effectiveness of numerical methods with the explainability and transferability of qualitative knowledge. While some progress in this area has already been made (e.g. transfer learning), hybrid forms of reasoning, combining qualitative and numerical forms of inference (e.g. inspired by dual process accounts of human reasoning) remain largely unexplored.

3.18 Uncertain Reasoning and Decision Theory

Wolfgang Spohn (Universität Konstanz, DE)

License  Creative Commons BY 3.0 Unported license
© Wolfgang Spohn

The talk modestly gives a survey of the various attempts of uncertainty measures and their extension to a decision theory. It will discuss the various theoretical achievements such an attempt has to provide, and it will emphasize the requirements such an attempt has to meet in order to be able to serve as a normative and/or as an empirical theory.

3.19 Knowledge, Uncertainty, and Ignorance

Sara L. Uckelman (*Durham University, GB*)

License © Creative Commons BY 3.0 Unported license
© Sara L. Uckelman

What is the relationship between knowledge, uncertainty, and ignorance? If knowledge is lack of uncertainty, and lack of knowledge is ignorance, then are ignorance and uncertainty the same thing? Drawing on arguments made by Paul of Venice (c1399) in his *Logica Magna*, we argue that they are ignorance and uncertainty are not the same thing; make a distinction between *mere* uncertainty and *fixed* uncertainty; and show how maybe knowledge shouldn't be defined as lack of uncertainty in the first place.

3.20 Possible Worlds Semantics for Conditionals: The Case of Chellas-Segerberg Semantics

Matthias Unterhuber (*LMU München, DE*)

License © Creative Commons BY 3.0 Unported license
© Matthias Unterhuber

This talk focuses on Chellas-Segerberg semantics, a base possible worlds semantics for conditionals. It is sketched in which way this semantics allows for structural frame conditions and can be expanded to a lattice of frames which allows one to describe a corresponding lattice of conditional logic system as described by thirty pairs of conditional logic principles and frame conditions.

In particular, it is explained which type of correspondence properties these pairs enjoy and which type of structure is required in order to arrive at a general non-trivial completeness result.

3.21 Knowledge and gossip

Hans Van Ditmarsch (*LORIA – Nancy, FR*)

License © Creative Commons BY 3.0 Unported license
© Hans Van Ditmarsch

Main reference M. Attamah, H. van Ditmarsch, D. Grossi, W. van der Hoek, “The Pleasure of Gossip,” in C. Baškent, L. Moss, R. Ramanujam (Eds.), “Rohit Parikh on Logic, Language and Society,” Springer, to appear; pre-print available at author's webpage.

URL <http://personal.us.es/hvd/FrohitFest.pdf>

Gossip protocols are to disseminate secrets by peer-to-peer communication in networks. In epistemic protocols the agents themselves choose whom to call. We present an example, and a version also involving exchange of telephone numbers.

3.22 Back to the Future – On the State of the Art in Default Reasoning

Emil Weydert (University of Luxembourg, LU)

License © Creative Commons BY 3.0 Unported license

© Emil Weydert

URL <http://materials.dagstuhl.de/files/15/15221/15221.EmilWeydert.Slides.pdf>

Default inference based on ranking measures (quasi-probabilistic plausibility valuations generalizing Spohn’s ranking functions, rational/real-valued) constitutes a powerful semantic approach to default reasoning. The idea is to let defaults induce constraints over ranking measures, to specify among the resulting ranking models preferred ones, and to use these to determine the defeasible conclusions conditional on a fact base. If we focus on those ranking models constructible by iterated revision with material implications reflecting the default base, which amounts to add a ranking weight for each default a world violates, we obtain well-behaved default consequence notions with nice inheritance features:

- System J (all the constructible models are preferred) – simple but rather weak.
- System ME (canonical preferred ranking model based on maximum entropy for non-standard probability) – probabilistic justification but representation-dependent, i.e. not invariant under boolean automorphisms, and not easy to compute.
- System JZ (canonical ranking construction implementing plausibility maximization, construction minimization, and justifiable constructibility).
- Z-style algorithm, verifies most desiderata, representation-independent.

Note however that for inheritance-friendly default formalisms, default bases are not characterized by their ranking-semantic content.

3.23 Dilation and Delayed Decisions

Gregory Wheeler (LMU München, DE)

License © Creative Commons BY 3.0 Unported license

© Gregory Wheeler

Joint work of Pedersen, Arthur Paul; Wheeler, Gregory

Main reference A. P. Pedersen, G. Wheeler, “Dilation, Disintegrations, and Delayed Decisions,” in Proc. of the 9th Int’l Symp. on Imprecise Probability: Theories and Applications (ISIPTA’15), pp. 227–236, 2015.

URL <http://www.sipta.org/isipta15/data/paper/23.pdf>

Dilation has been alleged to conflict with a fundamental principle of Bayesian methodology that we call Good’s Principle: one should always delay making a terminal decision between alternative courses of action if given the opportunity to first learn, at zero cost, the outcome of an experiment relevant to the decision. In particular, dilation has been alleged to permit or even mandate choosing to make a terminal decision in deliberate ignorance of relevant, cost-free information. This article presents dilation and a decision problem in which Good’s principle is violated. Our analysis shows that dilation, alone, is not enough to generate a violation of Good’s principle, but that the principle is only violated with respect to some decision rules, such as Gamma- Maximin, but not in terms of others, such as E-admissibility.

The slides also include a characterization result of dilation (which was not discussed in the talk) in terms of deviations from stochastic independence, which is a new result. The result tells us that dilation occurs when there are probability distributions in your “credal set” which render the two variables positively correlated and negatively correlated, which means that uncertainty about how the two variables are related to one another is a key feature

to dilation. We argue that in some circumstances discovering the possible consequences of your uncertainty concerning how the two variables are related to one another can be useful information to the decision-maker, meaning that dilation itself should not be viewed as a pathological feature of imprecise probability.

3.24 Representation and Bayesian Rational Predication

Momme von Sydow (Universität Heidelberg, DE)

License © Creative Commons BY 3.0 Unported license
© Momme von Sydow

Main reference M. von Sydow, “The Bayesian Logic of Frequency-Based Conjunction Fallacies,” *Journal of Mathematical Psychology*, 55(2):119–139, 2010.

URL <http://dx.doi.org/10.1016/j.jmp.2010.12.001>

In some contexts humans are able to deal with standard probabilities or even probability bounds. However, there are conceptual and empirical problems, if standard probability is used as a criterion of adequate predication. Whereas standard probabilities can directly be used for specifying the proportion of elements falling into a logically defined set, it is argued that for the goal of describing a situation in terms of logically connected predicates, standard probabilities do not (directly) provide a reasonable adequacy criterion: The Lockean thesis always allows to predicate more general but less informative logical hypotheses as well. Bayesian (pattern) logic (BL) addresses this problem by specifying the probability of alternative generative hypotheses (probability tables corresponding to standard truth tables) that provide a noisy-logical explanation or characterization of a situation. Here BL is not investigated in the context of dyadic logic, but in the context of monadic logic. We here extend this discussion to polytomous classes. BL predicts that the number of subclasses within an affirmation or within its negation should matter. The reported experiment provides evidence for this and shows strong deviations from standard probability (and also, for instance, from support theory). Although BL builds on standard extensional probabilities, it provides an intensional approach sensitive for the number of involved subclasses. Thus BL extends the scope of a probabilistic approach by advocating a goal-dependent pluralism within this approach.

4 Working Groups

4.1 Working Group “Combining and comparing qualitative and quantitative approaches to decision theory”

Ofer Arieli, Christoph Beierle, Tarik Besold, Florence Bannay-Dupin de Saint-Cyr, Steven Schockaert, Wolfgang Spohn, Sara L. Uckelman, and Emil Weydert

License © Creative Commons BY 3.0 Unported license
© Ofer Arieli, Christoph Beierle, Tarik Besold, Florence Bannay-Dupin de Saint-Cyr, Steven Schockaert, Wolfgang Spohn, Sara L. Uckelman, and Emil Weydert

URL <http://materials.dagstuhl.de/files/15/15221/15221.FlorenceBannay-Dupin%20de%20St-Cyr1.Slides2.pdf>

The result of the discussions have allowed us to propose a roadmap and define the challenges for the future.

4.1.1 Landscape

	Beliefs	Utility	Expected Utility
binary	B_1 B'_1	U_1 U'_1	E_1 E'_1
ordinary	B_2 B'_2	U_2 U'_2	E_2 E'_2
cardinal	B_3 B'_3	U_3 U'_3	E_3 E'_3

With $X \in \{B, U, E\}$, index 1=binary, 2=ordinary, 3=cardinal, X_i =complete knowledge, X'_i =vagueness/incompleteness

- comparison, partial or total completion (X'_i to X_i)
- comparison, combination X_i and X_j
- combination of B_i and U_j into E_k (and with ').

4.1.1.1 Examples

- B_3 : probability theory, U_3 : utility theory
Combination = Standard expected utility theory
- B_2 or B_3 : possibility theory U_3 : (standard) utility theory
Combination = Choquet (quantif) or Sugeno integral assumption “commensurability of possibility and utilities degrees”, axioms defined, the combination with Sugeno integral is the only one satisfying the axioms
NB: Choquet and Sugeno can use any uncertainty measure
- B_3 : General “Ranking theories” (e.g. Spohn) U_3 : (coarse-grained) ranking utility
Combination= ranking expected utility without coarse-grained perspective, no commensurability, may entail problems...
- B_3 or U_3 : 3 valuation classes: prob, rk, hybrid;
Combination across classes tricky, maybe hybrid B'_3 : convex sets of probabilities
- B_1 Epistemic and doxastic logic, U_1 : Conditional Deontic logic, logics of desires, goals (typically weak)
Combination C_1 : Some combined logics
- B_2 : total/partial plausibility orders, qualitative probability, U_2 : qualitative desire orders, theory of revealed preferences
(combinations of those? not much), C_2 Work in qualitative decision theory
- $B_1' = \dots?$, $U'_1 = (+, -, ?)$, $E'_1 = (+, -, ?)$

4.1.2 Dynamics

{Beliefs, Utilities, Expected utilities} can change

- revision
- update

4.1.3 Challenges

4.1.3.1 Theoretical Challenges

- What are the possible/sensible/applicable combinations?

- Axioms for each Box, and each combination (justified in a normative way and by experimentations) in order to ensure/evaluate the quality of the decision
 - ex:** Axiom on existence of independence notion
 - ex:** Axiom incompleteness of ordinals on convex or not convex sets
 - ex:** Axiom of comparability
- Imperfect rationality
 - inconsistency (beliefs, utilities, expected utilities)
 - bounded rationality about (impossible beliefs, utilities) or because bounded combination operator
 - ex:** Unawareness on utility values: “transformative experiences” ([a] la L.A. Paul) e.g. I am deciding whether to have children or not.
- How to build the beliefs? clarify the notion of beliefs.
- How to build the utilities (norms, desires...)
- If Input= Expected utility and Beliefs, how to reason?
- Strategical planning (theory of dynamic choice)= incomplete information planning with epistemic and ontic actions.
 - ex:** goal = to increase beliefs in order to become more expert for taking a better decision

4.1.3.2 Challenges from an Application Scenario

- (a) Express different kinds of uncertain, vague, incomplete knowledge coming from different sources, e.g.:
 - Probabilistic rules put forward by experts, taken from textbooks, like: “If symptoms A , B , and C are present, then D is the case with probability 0.8”.
 - Qualitative rules like: “If A and B , then D is more likely than D' ”.
 - Preferences like: ”For adults with biological age up to 50, prefer therapy type T to T' ”
- (b) Make inferences and suggest decisions:
 - Given some evidence, what is the most probable/likely diagnosis?
 - Which additional tests could be taken to increase diagnosis quality?
 - What are the options for a therapy plan?
 - What is the most sensible therapy given the current evidence? What is the risk for complications?
- (c) Challenges
 - Provide an adequate framework where all this can be expressed.
 - Ensure and show that inferences and (suggested) decisions are rational and justified.

4.2 Working Group “Probability & Inconsistency”

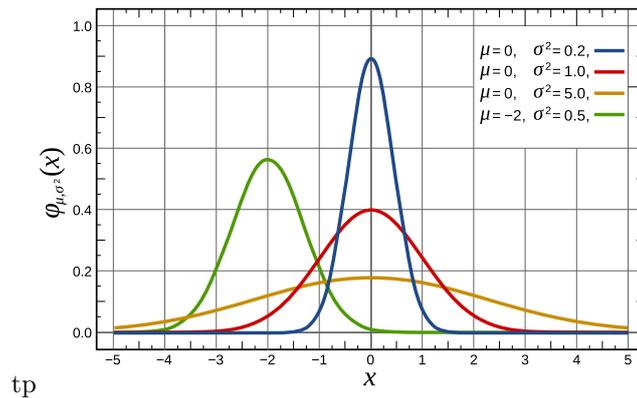
Nicole Cruz, Ulrike Hahn, Stephan Hartmann, Karolina Krzyżanowska, Momme von Sydow, Matthias Unterhuber, and Greg Wheeler

License © Creative Commons BY 3.0 Unported license

© Nicole Cruz, Ulrike Hahn, Stephan Hartmann, Karolina Krzyżanowska, Momme von Sydow, Matthias Unterhuber, and Greg Wheeler

URL <http://materials.dagstuhl.de/files/15/15221/15221.FlorenceBannay-Dupin%20de%20St-Cyr1.Slides2.pdf>

The result of the discussions have allowed us to propose a roadmap and define the challenges for the future.



■ **Figure 1** Normalized Gaussian curves with expected value μ and variance σ^2 (from: Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Gaussian_function&oldid=670515617)

4.2.1 Is the explosion problem a problem for the probabilist?

$A, \bar{A} \rightarrow \Phi \rightarrow P(\Phi|A \wedge \bar{A}) = 1$ But: $P(A \wedge \bar{A}) = 0$, so $P(\Phi|A \wedge \bar{A})$ undefined

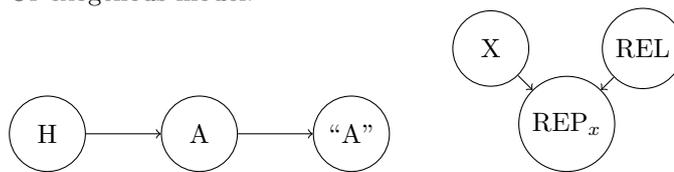
Explosion and its consequences cannot be expressed at the level of belief, so actually do not arise! (... even if consequences proliferate at level of logic). Nor is this possible in alternatives to Kolmogorov axioms such as Popper-Renyi functions etc. So answer ist “no”.

4.2.2 But what about data-bases etc.?

- Bayesian approach: consider A and \bar{A} to be tokens, outputs of a random variable
- *No inconsistency*: can reason effectively with this conflicting evidence, which is weighed and aggregated just like other evidence.

Though you do need a *model* ...

- $P("A"|H) \dots P("A"|H) \dots$ etc
- Or exogenous model:



Many standard models exist ...

- E.g. Figure 1
- Beyond Bayesianism (but within probabilities), non-parametric statistics ...

4.2.3 What about probalistic incoherence?

- Location of intuition that people hold inconsistent beliefs across different areas of their belief system: e.g., views on the economy that are incompatible with their views on politics (if they bothered to think it all through ...)

Incoherent systems of beliefs

- Incoherent beliefs seem bad, but at the same time, people seem to function well (and better than extant machine systems) in a noisy, changing world.

Pearl (1988)

- Fundamental claim that Bayesian computations are tractable in many contexts because what matters are comparatively local sets of conditional independence relations.
- Bayesian Networks encode and exploit these, and eliminate need to consider majority of joint probability distribution across variables.

Probabilistic incoherence across networks

- Are people's networks local in the sense of not being inter-linked?
- Or are they effectively local due to limited, weak links?
- Does it matter?

- Incoherence likely due to resource limitations in updating etc ...
- Enforcing coherence requires major collective effort: see e.g., law, physics, and relating variables across levels of description
- Is it worth it for bounded resource cognitive agents? Probably not, most of the time.
- Locality as a recipe for success?

“Probabilistic fault tolerance”

- How bad (empirically) is probabilistic incoherence?
- Do graphical models show graceful degradation with increasing incoherence?
- ... address through simulation (deliverable?!)
- Coarse graining: e.g., finite precision/rounding

Some limitations

- Poor fit graphical models?
 - Numerical values assigned to variables which are incoherent, i.e., there is no probability distribution which satisfies those values and the d -separation properties of a graph
- What to do outside the realm of Gaussian noise?
 - From Bayesian nets (single probabilities) to credal nets (sets of probabilities).

4.2.4 Abandoning models due to evidential conflict

- Hierarchical Bayesian approach (Bayesian inference about model selection e.g., Tenenbaum and colleagues..)
- Does this work in practice? Other approaches (probabilistic and non-probabilistic)

4.3 Working Group “Empirical Implications of formal systems for reasoning and vice versa”

Igor Douven, Christian Eichhorn, Thomas Eiter, Castaneda Gazzo, Estefania Lupita, Vittorio Girotto, Andreas Herzig, Anthony Hunter, Gabriele Kern-Isberner, Markus Knauff, Henry Markovits, Odile Papini, Niki Pfeifer, Henri Prade, Marco Ragni, Hans Rott, Henrik Singmann, and Hans Van Ditmarsch

License © Creative Commons BY 3.0 Unported license
 © Igor Douven, Christian Eichhorn, Thomas Eiter, Castaneda Gazzo, Estefania Lupita, Vittorio Girotto, Andreas Herzig, Anthony Hunter, Gabriele Kern-Isberner, Markus Knauff, Henry Markovits, Odile Papini, Niki Pfeifer, Henri Prade, Marco Ragni, Hans Rott, Henrik Singmann, and Hans Van Ditmarsch

What is normative?

- Philosophy: Normative theories are theories about what one should/ought do or how one should/ought reason.
- Psychology: A standard of reference (potentially rational) against which performance is evaluated. (needed for defining errors)
- CS: Human behavior can be a norm for AI.
- We basically all agree to the first two definitions.
- Bounded Rationality: Empirical norms should take cognitive/social constraints into account.

Role of Normative Ideas

- Psychology:
 1. Building blocks of descriptive theories of reasoning.
 2. Provide new empirical hypotheses (framework dominates task selection).
 3. No role for normative ideas: we should simply describe behavior.
- Phil: Try to identify and justify norms of reasoning and action.
- CS: Normative Ideas help to develop theories/models and help to establish prove properties of this models.

Non-monotonic logics

- Situated between two extremes: probability theory and classical logic.
- Try to remain connected to classical logic while including the idea of conditioning on the current state of knowledge.
- How can we empirically decide between whether or not individuals reason based on probability or using NM-logic. How can this influence psychological theories? How can this influence AI models?

Example: Rational Monotony. Rational monotony holds if for all A, B, C:

$$\frac{A \sim B \quad A \not\sim \neg C}{A \wedge C \sim B}$$

with

\sim : it normally follows, $\not\sim$: it normally does not follow

Example: Rational Monotony Violation. The following gives an example in which we expect rational monotony to not hold (i.e., we expect participants to disagree with the conclusion):

students \sim love reading books
 students $\not\sim$ do not love sports
 students who love sports \sim love reading books

5 Open Problems

From the discussions some central open questions could be identified.

First, clear definitions of concepts and ideas that all involved disciplines can agree upon are needed. As usual with multidisciplinary work, sometimes identical terms mean totally different concepts as well as identical concepts are described by different terms in the disciplines. Here, a common language is needed to smooth interdisciplinary work.

Apart from this very general open problem we also encountered several specific topics which should be worked on:

- Given the landscape in Section 4.1.1, we see the need to clarify the different functions for belief, utility and expected utility in a way that is useful for multi-disciplinary work.
- Based on this formal clarification it can be examined which combination of belief, utility and expected utility functions is the “best” combination in a given context.

Also we saw that there are different kinds of “vagueness” (e.g. probabilistic, qualitative, preferential, ...), which lead to the following questions:

- How can we express the different kinds of vagueness without mixing things up?
- How can we decide, infer and diagnose about combinations of these?
- Can we set up a framework that covers all of the different kinds of vagueness?

Participants

- Ofer Arieli
Academic College of Tel Aviv, IL
- Florence Bannay-Dupin de
St-Cyr
Paul Sabatier University –
Toulouse, FR
- Christoph Beierle
FernUniversität in Hagen, DE
- Tarek R. Besold
Universität Osnabrück, DE
- Nicole Cruz de Echeverria
Loebell
University of London, GB
- Igor Douven
Paris-Sorbonne University, FR
- Christian Eichhorn
TU Dortmund, DE
- Thomas Eiter
TU Wien, AT
- Lupita Estefania Gazzo
Castaneda
Universität Gießen, DE
- Vittorio Giroto
University of Venezia, IT
- Ulrike Hahn
University of London, GB
- Stephan Hartmann
LMU München, DE
- Andreas Herzig
Paul Sabatier University –
Toulouse, FR
- Anthony Hunter
University College London, GB
- Gabriele Kern-Isberner
TU Dortmund, DE
- Markus Knauff
Universität Gießen, DE
- Karolina Krzyzanowska
LMU München, DE
- Henry Markovits
University of Montreal, CA
- Odile Papini
University of Marseille, FR
- Niki Pfeifer
LMU München, DE
- Henri Prade
Paul Sabatier University –
Toulouse, FR
- Marco Ragni
Universität Freiburg, DE
- Hans Rott
Universität Regensburg, DE
- Steven Schockaert
Cardiff University, GB
- Henrik Singmann
Universität Zürich, CH
- Wolfgang Spohn
Universität Konstanz, DE
- Sara L. Uckelman
Durham University, GB
- Matthias Unterhuber
LMU München, DE
- Hans Van Ditmarsch
LORIA – Nancy, FR
- Momme von Sydow
Universität Heidelberg, DE
- Emil Weydert
University of Luxembourg, LU
- Gregory Wheeler
LMU München, DE



Human-Centric Development of Software Tools

Edited by

Andrew J. Ko¹, Shriram Krishnamurthi², Gail Murphy³, and Janet Siegmund⁴

1 University of Washington, USA, ajko@uw.edu

2 Brown University, USA, sk@brown.edu

3 University of British Columbia, Canada, murphy@cs.ubc.ca

4 University of Passau, Germany, Janet.Siegmund@uni-passau.de

Abstract

Over two and half days, over 30 participants engaged in inventing and evaluating programming and software engineering tools from a human rather than tool perspective. We discussed methods, theories, recruitment, research questions, and community issues such as methods training and reviewing. This report is a summary of the key insights generated in the workshop.

Seminar May 25–28, 2015 – <http://www.dagstuhl.de/15222>

1998 ACM Subject Classification H.5.2 User Interfaces, D.2.3 Coding Tools and Techniques

Keywords and phrases Software engineering, developer tools, human factors

Digital Object Identifier 10.4230/DagRep.5.5.115

1 Executive Summary

Andrew J. Ko

License  Creative Commons BY 3.0 Unported license
© Andrew J. Ko

Across our many sessions, we discussed many central issues related to research on the design of human-centric developer tools. In this summary, we discuss the key insights from each of these areas, and actionable next steps for maturing the field of human-centered developer tools.

Key Insights

Theories

Theories are a hugely important but underused aspect of our research. They help us start with an explanation, they help us explain and interpret the data we get, they help us relate our findings to others findings, and they give us vocabulary and concepts to help us organize our thinking about a phenomenon.

There are many relevant theories that we should be using:

- Attention investment is helpful in explaining why people choose to engage in programming.
- Information foraging theory helps explain where people choose to look for relevant information in code.
- Community of practice theory helps us explain how people choose to develop skills over time.



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

Human-Centric Development of Software Tools, *Dagstuhl Reports*, Vol. 5, Issue 5, pp. 115–132

Editors: Andrew J. Ko, Shriram Krishnamurthi, Gail C. Murphy, and Janet Siegmund



DAGSTUHL
REPORTS

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

There are useful methods for generating theories, including grounded theory and participatory design. Both can result in explanations of phenomena. That said, there are often already theories about things and we don't need to engage in creating our own.

While theories are the pinnacle of knowledge, there's plenty of room for "useful knowledge" that helps us ultimately create and refine better theories. Much of the research we do now generates this useful knowledge and will eventually lead to more useful theories.

Study Recruitment

Whether developers agree to participate in a study depends on several factors:

- One factor is how much value developers perceive in participating. Value might be tangible (a gift card, a bottle of champagne), or personal (learning something from participation, or getting to share their opinion about something they are passionate about).
- Another factor in recruitment is whether the requestor is part of the developer in-group (e.g, being part of their organization, having a representative from their community conduct the research or recruit on your behalf, become part of their community before asking for their efforts)
- The cost of participating obviously has to be low, or at least low enough to account for the benefit. With these factors in mind, there are a wide range of clever and effective ways to recruit participants:

Monitor for changes in bug databases and gather data at the moment the event occurs. This makes the request timely and minimizes the cost of recall.

- Find naturalistic captures of people doing software engineering work (such as tutorials, walkthroughs, and other recorded content that developers create). This costs the nothing.
- Perform self-ethnographies or diary studies. This has some validity issues, but provides a rich source of data.
- Tag your own development work through commits to gather interesting episodes.
- Find where developers are and interview them there (e.g., the Microsoft bus stop, developer conferences), and generate low-cost, high-value ways of getting their attention (and data).

Research Questions

There was much discussion of research questions at the conference and what makes a good one. There was much agreement that our questions should be more grounded in theories, so that we can better build upon each others' work.

Many researchers also find that the human-centered empirical studies produce results that are not directly meaningful or actionable to others. There are many possible reasons for this:

- We often don't choose research questions with more than one plausible outcome.
- We often don't report our results in a way that creates conflict and suspense. We need to show readers that there are many possible outcomes.
- We often ask "whether" questions, rather than "why" or "when" questions about tools, leading to limited, binary results, rather than richer, more subtle contributions.

Some of our research questions have validity issues that make them problematic:

- Research questions often fail to understand the populations they are asking about.
- Research questions often get involved in designing tools for people who are already designing tools for themselves. Instead, researchers should be building tools that have never existed, not building better versions of tools that already exist.

One opportunity for collaboration with researchers who are less human-centered is to collaborate on formative research that shapes the direction of research and discover new research opportunities for the field. This may create more positive perceptions of our skills, impact, and relevance to the broader fields of PL and SE.

Human-Centeredness

Historically, HCI concerns have focused on end user experiences rather than developer experiences, but HCI researchers have increasingly focused on developers and developer tools. But HCI often doesn't consider the culture and context of software engineering, and doesn't address the longitudinal / long term factors in education and skill acquisition, and so HCI may not be a sufficient lens through which to understand software engineering.

There is also a need to address low-end developers, not just "experts". Future research topics include the understand learnability of APIs, how to understand the experiences of engineers (from a sociological perspective studies such as Bucciarelli), how to think about tools from a knowledge prerequisite perspective.

Developer Knowledge Modeling

Much of what makes a developer effective is the knowledge in their mind, but we know little about what this knowledge is, how developers acquire it, how to measure and model it, and how to use these models to improve tools or enable new categories of tools. There are many open opportunities in this space that could lead to powerful new understandings about software engineering expertise and powerful new tools to support software engineering. Much of this new work can leverage research in education and learning sciences to get measures of knowledge.

Leveraging Software Development Analytics

We identified identifying different types of data that might be collected on programming processes and products. These included editing activities, compilation attempts and errors, execution attempts and errors, and check-ins. We considered ways in which these data could be enlisted to help improve teaching and learning, as well as the software development process:

- Automated interventions to improve programming processes
- Present visually to aid in decision making
- Generate notifications that could inform learners, teachers, and software developers of key events.
- Generating social recommendations.

These opportunities raise several questions:

- How do we leverage data to intervene in educational and collaborative software development settings?
- How do we design visual analytics environment to aid in decision making?
- Should interventions be automated, semi-automated, or manual? What are the trade offs?

Error Messages

We identified 5 broad classes of errors: (1) syntactic (conformance to a grammar), (2) type, (3) run-time (safety checks in a run-time system, such as array bounds, division by zero,

etc.), (4) semantic (logical errors that aren't run-time errors) (5) stylistic. We distinguished between errors and more general forms of feedback, acknowledging that both needed support; in particular, each of these could leverage some common presentation guidelines.

We discussed why research has tended to focus more on errors for beginners than feedback for developers. Issues raised included the different scales of problems to diagnose across the two cases and differences in social norms around asking for help from other people (developers might be less likely to ask other people for help in order to protect their professional reputations). We discussed whether tools should report all errors or just some of them, and whether tools should try to prioritize among errors when presenting them. These had different nuances in each of students and practicing developers. We discussed the example of the coverity tool presenting only a subset of errors, since presenting all of them might lead developers to reject the tool for finding too much fault in their code.

We discussed and articulated several principles of presenting errors: (1) use different visual patterns to distinguish different kinds of errors; (2) don't mislead users by giving incorrect advice on how to fix an error; (3) use multi-dimensional or multi-modal techniques to reveal error details incrementally; (4) when possible, allow programs to fail gently in the face of an error (example: soft typing moved type errors into run-time errors that only tripped when a concrete input triggered the error – this gives the programmer some control over when to engage with the error after it arises); (5) consider ways to allow the user to query the system to narrow down the cause of the error (rather than require them to debug the entire program).

There are several open research questions:

- Should error and feedback systems become interactive, asking the user questions to help diagnose a more concrete error (rather than report a more abstract one, as often happens with compiler syntax errors)?
- Can grammars be tailored to domain-specific knowledge to yield more descriptive error messages?
- Can patterns of variable names be used to enforce conventions and reduce the rates of some kinds of errors?
- At what point should error systems expect the user to consult with another human, rather than rely only on the computer.
- When is it more helpful to show all errors (assuming we can even compute that) versus a selection of errors? How much detail should be presented about an error at first? Does presenting all information discourage users from reading error messages?

Reviewing

Researchers in human aspects of software engineering feel a strong sense of hostility towards human-centered research, despite some recent successes in some software engineering venues. Reasons for this hostility include:

- Many human-centered researchers evaluate and critique tools without offering constructive directions forward. This creates a perception that human-centered researchers dislike or hate the research that others are doing.
- Many human-centered researchers are focused on producing understanding, whereas other researchers are focused on producing better tools. This goal mismatch causes reviewers to apply inappropriate criteria to the importance and value of research contributions.
- Many research communities in programming languages and software engineering still lack sufficient methodological expertise to properly evaluate human-centered empirical work.

- It's not explicit in reviews whether someone's methodological expertise is a good match for a paper. Expert in a topic, not expert in a method. This leads to topic expertise matches without methodological expertise matches.
- Many challenges in reviewing come from the difference between judging a paper's validity versus judging how interesting a paper is. Non-human centered researchers do not often find our questions interesting.

We are often our own worst enemies in reviews. We often reject each other because we're too rigid about methods (e.g., rejecting papers because of missing interrater reliability). On the other hand, we have to maintain standards. There's a lot of room for creativity in establishing rigor that is satisfying to reviewers, and we should allow for these creative ways of validating and verifying our interpretations.

Methods Training

Empirical methods are not popular to learn. However, when our students and colleagues decide to learn them, there are many papers, textbooks, classes and workshops for learning some basic concepts in human-subjects software engineering research.

There are many strategies we might employ to broadly increase methodological expertise in our research communities:

- We should spend more time in workshops and conferences teaching each other how to do methods well.
- Software engineers need to learn empirical methods too, and teaching them as undergraduates will lead to increased literacy in graduate students.
- There is much we can do to consolidate and share teaching resources that would make this instruction much more efficient.
- HCI research methods are broadly applicable and there are many more places to learn them.

There aren't good methods for researching learning issues yet. Moreover, most of these methods cannot be learned quickly. We must devise ways of teaching these methods to students and researchers over long periods of time.

2 Table of Contents

Executive Summary

<i>Andrew J. Ko</i>	115
-------------------------------	-----

Poster Abstracts

Teaching Usability of Programming Languages	
---	--

<i>Alan Blackwell</i>	122
---------------------------------	-----

The GenderMag Kit: To Find Usability Issues from a Gender Perspective	
---	--

<i>Margaret M. Burnett</i>	122
--------------------------------------	-----

Software Tools and Practices	
------------------------------	--

<i>Yvonne Dittrich</i>	123
----------------------------------	-----

Programming Languages as Interfaces in Plan Composition	
---	--

<i>Kathi Fisler</i>	125
-------------------------------	-----

Empirical Support for Contextual Computing Educaiton	
--	--

<i>Mark Guzdial</i>	125
-------------------------------	-----

Socio-Technical Coordination: How Millions of People use Transparency to Collaborate on Millions of Interdependent Projects on GitHub	
---	--

<i>James D. Herbsleb</i>	125
------------------------------------	-----

Increasing tool reach through online integration	
--	--

<i>Reid Holmes</i>	126
------------------------------	-----

Thoughts about Evidence-Based Programming Language Design	
---	--

<i>Antti-Juhani Kaijanaho</i>	126
---	-----

Explaining Software One Bit at a Time	
---------------------------------------	--

<i>Andrew J. Ko</i>	127
-------------------------------	-----

User Interfaces for Error Reporting	
-------------------------------------	--

<i>Shriram Krishnamurthi</i>	127
--	-----

Bespoke Tools: Adapted to the Concepts Developers Know	
--	--

<i>Emerson Murphy-Hill</i>	127
--------------------------------------	-----

Bridging between Research and Adoption of Software Tools	
--	--

<i>Gail C. Murphy</i>	128
---------------------------------	-----

Using the Natural Programming Approach Throughout the Lifecycle	
---	--

<i>Brad A. Myers</i>	128
--------------------------------	-----

Supporting Developers Decision	
--------------------------------	--

<i>Barbara Paech</i>	128
--------------------------------	-----

Identifying Barriers to Participation of Females Users on Stack Overflow	
--	--

<i>Chris Parnin</i>	129
-------------------------------	-----

The Forgetting Curve	
----------------------	--

<i>Peter C. Rigby</i>	129
---------------------------------	-----

Does UML Diagram Layout Affect Model Understanding?	
---	--

<i>Harald Stoerrle</i>	129
----------------------------------	-----

Programming in Natural Language	
---------------------------------	--

<i>Walter F. Tichy</i>	130
----------------------------------	-----

Industry adoption requires empirical evaluations in a real context
Claes Wohlin 130

Productivity and Data Science for Software Engineering
Thomas Zimmermann 131

Participants 132

3 Poster Abstracts

3.1 Teaching Usability of Programming Languages

Alan Blackwell (University of Cambridge, GB)

License © Creative Commons BY 3.0 Unported license
© Alan Blackwell

Main reference Usability of Programming Languages. Lecture notes for MPhil in Advanced Computer Science, University of Cambridge Computer Laboratory

URL <http://www.cl.cam.ac.uk/teaching/1415/P201/p201-lecturenotes-2015.pdf>

I presented an overview of my graduate course teaching usability of programming languages. The course itself is documented here: <http://www.cl.cam.ac.uk/teaching/1415/P201/>. Lecture notes for the course are in the bibliographic reference.

3.2 The GenderMag Kit: To Find Usability Issues from a Gender Perspective

Margaret M. Burnett (Oregon State University, US)

License © Creative Commons BY 3.0 Unported license
© Margaret M. Burnett

URL <http://eusesconsortium.org/gender>

The way people use software features often differs according to their gender, especially when the software is for problem-solving, as when figuring out budgets, or understanding visualizations, or debugging. However, many software features are inadvertently designed around the way males tend to use software.

How using GenderMag helps make software gender-inclusive:

The GenderMag method helps software developers and usability professionals identify features that don't take females' common usage patterns into account. The GenderMag method can also be used to find features that don't take males' common usage patterns into account.

- The GenderMag method encapsulates how five facets of gender differences (motivations for use, information processing style, computer self-efficacy, attitude toward risk, and willingness to explore/tinker) affect the ways males and females tend to use software.
- Software developers and usability professionals use the method to find gender-inclusiveness issues. They can then fix these issues one at a time, so as to individually take down barriers that may disproportionately affect one gender but could also affect a fraction of the other.
- Researchers who have used early versions of GenderMag in this way have found that their software becomes more gender-inclusive and better liked by its users overall.

GenderMag consists of a set of GenderMag Personas to bring the facets to life, and a GenderMag Cognitive Walkthrough to embed use of the personas in a process.

The GenderMag Personas

- Personas represent “archetypes” of users of a software system. The focus of the male and female GenderMag personas is bringing to life the above five facets of gender differences.

The GenderMag Cognitive Walkthrough

- The GenderMag CW is a gender-specialized Cognitive Walkthrough (CW). A (regular) CW analyzes how easily users new to the system can accomplish specific tasks with that system.
- The GenderMag CW adds to the analysis process explicit use of the above five facets of gender differences and the personas.

No background in gender difference research is needed to use GenderMag. It is intended for any software developer or usability professional interested in identifying the features of their software that may not be gender-inclusive.

3.3 Software Tools and Practices

Yvonne Dittrich (IT University of Copenhagen, DK)

License © Creative Commons BY 3.0 Unported license
© Yvonne Dittrich

Main reference Y. Dittrich, “What does it mean to use a method? Towards a practice theory for software engineering,” *Information and Software Technology*, 70(Feb. 2015):220–231, 2015.

URL <http://dx.doi.org/10.1016/j.infsof.2015.07.001>

How tools are used depends on how they are embedded in specific practices. Tools often take specific processes and methods for given. However, as we all know, the use of the same tool does not necessarily result in comparable impacts on the software development practice. E.g. Damian et al. (2007) report on a case of a team distributed between US and Canada, using the same method and tooling. One issue was the use of the Code Versioning System. Whereas in one locality, the team members relied on the commit comments distributed to the whole team, the team members in the other locality used extra mails distributed via a mailing list to highlight changes affecting other members of the team. Damian et al. attributed the resulting break-downs to cultural differences. Culture here refers to locally shared but across sides diverging practices. (See also Agar 1996, p. 9). But what are practices?

A Practice Concept for Software Engineering

In philosophy, the notion of practice has been discussed since the time of ancient Greek philosophy. A historical overview of the development of the practice concept is found in (Schmidt 2014). Schmidt defines practice as ‘normative regulated contingent activity’ (Schmidt 2014, p. 437), arguing that the modern concept of practice focuses ‘on the ways in which the competent actor in his or her action is taking the particular conditions into account while committed to and guided by the appropriate general principles (‘theory’, ‘rules’)’ (Schmidt 2014, p. 436). Wittgenstein’s *Philosophical Investigations* have inspired a number of contemporary schools in social science and philosophy. In his introduction to practice theory, Nicolini (2012) talks about practice theories. To ground a practice theory in a coherent and consistent manner, I refer in (Dittrich 2015) to Schatzki’s ‘Social Practice. A Wittgensteinian approach to human activity and the social’ (1996) as well as Knorr Cetina’s article ‘Objectual Practice’ (2001): Schatzki defines, based on Wittgenstein, practice as a “. . . temporally unfolding and spatially dispersed nexus of doings and sayings. (. . .) to say that the doings and sayings forming a nexus is to say that they are linked in certain ways. Three major avenues of linkage are involved: 1) through understandings, for example, of what to say and do; (2) through explicit rules, principles, precepts, and instructions;

and (3) through what I will call teleoaffective structures embracing ends, projects, tasks, purposes, beliefs, emotions, and moods.” [12, p. 89] Such practices take place in specific settings making use of tools, materials and objects that acquire their meaning through the practice they are supporting (Schatzki 1996, p. 113-114). Based on Knorr Cetina’s concept of epistemic practices (2001), I argue that software development is an epistemic practice that unfolds its object and, with it, its own practice as the team proceeds in the development. Such adaptation of practices has been observed as meta and articulation work (Sigfridsson 2010) and as joint tailoring of development environments (Draxler et al. 2014) in software engineering projects. Based on this foundation, methods can be defined as practice patterns, explicitly formulated sets of (tool supported) understandings, rules and teleoaffective structures that need to be integrated in existing practices. See (Dittrich 2015) for a detailed development of the argument.

What does that imply for the use and the usefulness of tools? The usability and usefulness of a tool depends on how it can be embedded into concrete practices. The adaptation of tools can be observed as a conscious and continuous effort of software engineering teams (Draxler et al. 2014; Giuffrida and Dittrich 2015).

Software engineering practices, however, are situated with respect to organisation, domain, and even the specific project. They are not stable but change with the developing understanding of the software product they are meant to develop or evolve. The introduction of a tool implies a development of the practices it is meant to support and needs to be accompanied by meta-work (Strauss 1985), explicit negotiation and agreement of how to embed the tool in the practice. That means that the transfer of tools that are developed based on scientific ideas of programming techniques and languages needs to explicitly consider how it relates to industrial practices. Empirical research should focus on the embedment of the new tool into exemplary practices.

References

- 1 M. H. Agar, *The professional stranger: An informal introduction to ethnography*. San Diego, CA: Academic Press, 1996.
- 2 D. Damian, L. Izquierdo, J. Singer, I. Kwan, Awareness in the wild: Why communication breakdowns occur. In *International Conference on Global Software Engineering (ICGSE) 2007*.
- 3 S. Draxler, G. Stevens, A. Boden, Keeping the development environment up to date-A Study of the Situated Practices of Appropriating the Eclipse IDE. *IEEE Transaction on Software Engineering*, 40 (2014), pp. 1061–1074.
- 4 Y. Dittrich, What does it mean to use a method? Towards a practice theory for software engineering. *Information and Software Technology*, 2015 in print.
- 5 R. Giuffrida, Y. Dittrich, A conceptual framework to study the role of communication through social software for coordination in globally-distributed software teams. *Information and Software Technology*, 63 (2015), pp. 10–30.
- 6 K. Knorr Cetina, Objectual practice. In: *The practice turn in contemporary theory*, T. Schatzki, K. Knorr Cetina, and E. von Savigny (eds.), London: Routledge 2001, pp. 175–188.
- 7 D. Nicolini, *Practice theory, work, and organization: an introduction*. Oxford University Press, Oxford 2012.
- 8 T.R. Schatzki, *Social practices: A Wittgensteinian approach to human activity and the social*. Cambridge University Press 1996.
- 9 K. Schmidt, The Concept of Practice: What’s the Point?. In *COOP 2014-Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27-30 May 2014, Nice (France)*, Springer Intern. Publ., pp. 427–444.

- 10 A. Sigfridsson, The purposeful adaptation of practice: an empirical study of distributed software development. PhD Thesis. University of Limerick 2010.
- 11 A.L. Strauss, Work and the division of labor. *Sociol Q*, 26(1985), pp. 1–19.
- 12 L. Wittgenstein, *Philosophical Investigations*. The German text, with an English translation by G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte. Revised 4th edition by P.M.S. Hacker and Joachim Schulte. Wiley-Blackwell 2009.

3.4 Programming Languages as Interfaces in Plan Composition

Kathi Fisler (Worcester Polytechnic Institute, US)

License  Creative Commons BY 3.0 Unported license
© Kathi Fisler

This poster explored the ways in which the language constructs that students know affect how they structure solutions to programming problems. The relationship between constructs and solution structure has pedagogic implications: the structures we want students to produce would seem to dictate either the constructs we teach or the problems we assign. This has human-factors implications since some solutions are harder to implement correctly than others. We illustrate the issues with two concrete problems that we have been using in educational studies this year.

3.5 Empirical Support for Contextual Computing Educaiton

Mark Guzdial (Georgia Institute of Technology – Atlanta, US)

License  Creative Commons BY 3.0 Unported license
© Mark Guzdial

Programming is the way it is because of skewed demographics. Changing who programs will dramatically change programming. Changing the language or IDE will do little to change who programs. Humans care about narrative, the context around the activity. People reject the context of programming, not the activity of programming. If we teach programming within context, we broaden the range of people who program.

3.6 Socio-Technical Coordination: How Millions of People use Transparency to Collaborate on Millions of Interdependent Projects on GitHub

James D. Herbsleb (Carnegie Mellon University, US)

License  Creative Commons BY 3.0 Unported license
© James D. Herbsleb

Joint work of Dabbish, Laura; Stuart, Colleen; Tsay, Jason; Herbsleb, James D.

Main reference L. Dabbish, C. Stuart, J. Tsay, J. Herbsleb, “Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository,” in *Proc. of the ACM 2012 Conf. on Computer Supported Cooperative Work (CSCW’12)*, pp. 1277–1286, ACM, 2012.

URL <http://dx.doi.org/10.1145/2145204.2145396>

Collectively creating digital things these days often means hoards of people collaborating in open, transparent environments, loosely organized in ecosystems of interdependent projects.

Splitting work across large collections of people has great potential benefits such as tapping a larger talent pool, enabling better matches between tasks and skills, and reducing schedule bottlenecks. But it also gives rise to difficult coordination problems while disabling coordination mechanisms that rely on overarching hierarchies of authority. In this talk, I will develop a socio-technical theory of coordination, and show how colleagues and I empirically validated it in a geographically distributed software development organization. I will show how the theory can be adapted to help interpret the results our qualitative study of coordination practices in GitHub, an open, transparent work environment in which millions of people collaborate on millions of interdependent projects.

3.7 Increasing tool reach through online integration

Reid Holmes (University of Waterloo, CA)

License © Creative Commons BY 3.0 Unported license
© Reid Holmes

Many development tools have high-quality web interfaces. Software engineering tools can greatly increase their reach by integrating with these interfaces, rather than forcing developers to download and install new tools or to change their development processes by visiting new sites. By injecting content into existing web interfaces, new development tools can be seamlessly integrated into the developer's current tools, increasing the tool's ability to augment and improve the developer's current experience without disrupting the way they work.

3.8 Thoughts about Evidence-Based Programming Language Design

Antti-Juhani Kaijanaho (University of Jyväskylä, FI)

License © Creative Commons BY 3.0 Unported license
© Antti-Juhani Kaijanaho

Main reference A.-J. Kaijanaho, "The extent of empirical evidence that could inform evidence-based design of programming languages: a systematic mapping study," Jyväskylä Licentiate Theses in Computing 18, University of Jyväskylä, 2014.

URL <http://urn.fi/URN:ISBN:978-951-39-5791-9>

In this poster, I presented highlights from my systematic mapping study on the extent of empirical evidence that could inform evidence-based programming language design. I particularly pointed out the curious pattern of publications in this area, where publication seems to have moved from HCI forums to technical programming language forums, and a clear upsurge in publications has occurred in recent years.

In addition, I also discussed my currently unpublished thoughts on empirical evaluation of language design choices, as well as my explication of Evidence-Based Programming Language Design, which will be a significant part of my doctoral dissertation, to be submitted very soon.

3.9 Explaining Software One Bit at a Time

Andrew J. Ko (University of Washington – Seattle, US)

License © Creative Commons BY 3.0 Unported license
© Andrew J. Ko

Modern software is increasingly complex, making it ever more difficult to use, understand, and fix. The USE research group invents technologies that help people understand and overcome this complexity, including new help systems for end users, new debugging tools for developers, and new educational technologies for people learning to program. Additionally, we conduct a wide range of studies about software engineering teams, programming expertise, and learning. Our work spans human-computer interaction, software engineering, and computing education.

3.10 User Interfaces for Error Reporting

Shriram Krishnamurthi (Brown University – Providence, US)

License © Creative Commons BY 3.0 Unported license
© Shriram Krishnamurthi

Joint work of Marceau, Guillaume; Fisler, Kathryn; Krishnamurthi, Shriram

Main reference G. Marceau, K. Fisler, S. Krishnamurthi, “Measuring the Effectiveness of Error Messages Designed for Novice Programmers,” in Proc. of the 42nd Technical Symp. on Computer Science Education (SIGCSE’11), pp. 499-504, ACM, 2011.

URL <http://dx.doi.org/10.1145/1953163.1953308>

Errors are a critical, inescapable part of programming. How errors are reported therefore has an impact on how programmers perceive the programming experience. This is especially likely to impact beginning programmers, who do not have the experience to know what to do, and are more likely to internalize these mistakes as statements about themselves (with a corresponding impact on their confidence).

Over the past few years, Marceau, Fisler, and I have been studying the quality of error messages in programming environments, and identifying ways to improve them. We have prototyped some new interfaces, which we are now setting up to study.

3.11 Bespoke Tools: Adapted to the Concepts Developers Know

Emerson Murphy-Hill (North Carolina State University – Raleigh, US)

License © Creative Commons BY 3.0 Unported license
© Emerson Murphy-Hill

Main reference B. Johnson, R. Pandita, E. Murphy-Hill, S. Heckman, “Bespoke tools: adapted to the concepts developers know,” in Proc. of the 2015 10th Joint Meeting on Foundations of Software Engineering (ESEC/FSE’15), pp. 878–881, ACM, 2015.

URL <http://dx.doi.org/10.1145/2786805.2803197>

Even though different developers have varying levels of expertise, the tools in one developer’s integrated development environment (IDE) behave the same as the tools in every other developers’ IDE. In this paper, we propose the idea of automatically customizing development tools by modeling what a developer knows about software concepts. We then sketch three such “bespoke” tools and describe how development data can be used to infer what a developer knows about relevant concepts. Finally, we describe our ongoing efforts to make bespoke program analysis tools that customize their notifications to the developer using them

3.12 Bridging between Research and Adoption of Software Tools

Gail C. Murphy (University of British Columbia – Vancouver, CA)

License © Creative Commons BY 3.0 Unported license
© Gail C. Murphy

After the hard and creative work of researching, defining and building a tool to aid software development comes the even harder parts: what does it take to get your tool adopted into wide-scale use. Engineering the tool to be easy to use, deployable, scalable, etc. takes some work. But even more work is needed to truly understand the users base, make the tool apparent to use, make the tool fit seamlessly into software development workflows, make the tool handle the multitude of different environments developers use and so on. In the work presented at Dagstuhl, I discussed a number of challenges we faced, and largely overcame, in having a research tool, Mylyn, adopted into wide-scale use and subsequently commercialized in a different form at Tasktop Technologies.

3.13 Using the Natural Programming Approach Throughout the Lifecycle

Brad A. Myers (Carnegie Mellon University, US)

License © Creative Commons BY 3.0 Unported license
© Brad A. Myers
URL <http://www.cs.cmu.edu/~NatProg/>

- Make programming easier by making it more natural, by which we mean closer to the way people think about their tasks.
- Apply in all phases of tool development.
- Use a large variety of HCI methods.

3.14 Supporting Developers Decision

Barbara Paech (Universität Heidelberg, DE)

License © Creative Commons BY 3.0 Unported license
© Barbara Paech

Developers make decisions about the software to be built and the software engineering process. To make adequate decisions about the software they need to understand the users and the former decisions of developers (e.g. about the architecture). The tool UNICASE (www.unicase.org) supports capturing and navigating through the system knowledge (requirements, design, code, test) and project knowledge (e.g. work items) as well as the decisions and their rationale. To understand how developers make and document decisions we have made a study (which is to be submitted shortly) on decision knowledge and decision strategies in issue trackers. This study confirms the importance of naturalistic-decision making. Therefore, in UNICASE decision knowledge can be captured incrementally evolving from a naturalistic decision to a rationale decision with thorough criteria and arguments. In the project URES (<http://www.dfg-spp1593.de/index.php?id=38>) we study in addition the capture of software usage knowledge. The interaction of the user with the software is monitored on a high-level, e.g. to compare it with use cases in the system knowledge and then improve the the use cases according to actual usage.

3.15 Identifying Barriers to Participation of Females Users on Stack Overflow

Chris Parnin (North Carolina State University – Raleigh, US)

License © Creative Commons BY 3.0 Unported license
© Chris Parnin

Joint work of Parnin, Chris; Ford, Denae; Smith, Justin; Guo, Philip

It is no secret that females participate in the programming field less than males. This gender gap is also evident on Stack Overflow where in a recent survey only 5.8% of 26,086 respondents identified as female. This study aims to help understand low participation of females on Stack Overflow. Through a manual inspection of users profiles, we found only a 0.25% female participation rate in the top 108,000 user accounts. Through 22 semi-structured interviews with general female users, including an interview with a female user currently ranked as one of the top 100 Stack Overflow users, we identify several barriers, gender-related as well as general barriers specific to Stack Overflow’s design. This paper explains why females do not use the site to it’s full potential and provides interventions as to how to encourage them to. We found they did not participate, not only because of the time it would take to use the site, but also because they face pressures to research their posts, resulting in them choosing to ask less questions at a lower rate than males.

3.16 The Forgetting Curve

Peter C. Rigby (Concordia University – Montreal, CA)

License © Creative Commons BY 3.0 Unported license
© Peter C. Rigby

There has been a great deal of research looking into how quickly students forget what they have been taught. The goal of this research is to investigate how much developers remember of the code they’ve written. This will help us understand how much knowledge about the system the development team actually has. It will also allow us to assess how long creative professionals retain information about what they’ve created.

3.17 Does UML Diagram Layout Affect Model Understanding?

Harald Storrle (Technical University of Denmark – Lyngby, DK)

License © Creative Commons BY 3.0 Unported license
© Harald Storrle

Diagrams are widely used in Software Engineering. Intuitively, good layouts are very helpful when understanding UML diagrams. However, existing studies were inconclusive. So, does layout matter?

In a series of experiments we found evidence that layouts do actually matter. We have studied individual factors and the underlying cognitive mechanisms with a view to improving notations and diagramming practice.

3.18 Programming in Natural Language

Walter F. Tichy (KIT – Karlsruher Institut für Technologie, DE)

License © Creative Commons BY 3.0 Unported license
© Walter F. Tichy

Main reference M. Landhäuser, T. Hey, W. Tichy, “Deriving Time Lines from Texts,” in Proc. of the 3rd Int’l Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE’14), pp. 45–51, ACM, 2014.

URL <http://dx.doi.org/10.1145/2593801.2593809>

Natural language interfaces are becoming more and more common, because they are powerful and easy to use. Examples of such interfaces are voice controlled navigation devices, Apple’s personal assistant Siri, Google Voice Search, and translation services. However, such interfaces are extremely challenging to build, maintain, and port to new domains.

We present an approach for building and porting such interfaces quickly. NLCI is a natural language command interpreter that accepts action commands in English and translates them into executable code. The core component is an ontology that models an API. Once the API is “ontologized”, NLCI translates input sentences into sequences of API calls that implement the intended actions. Two radically different APIs were ontologized: openHAB for home automation and Alice for building 3D animations. Construction of the ontology can be automated if the API uses descriptive names for its components. In that case, the language interface can be generated completely automatically.

Recall and precision of NLCI on a benchmark of 50 input scripts are 67% and 78 %, resp. Though not yet acceptable for practical use, the results indicate that the approach is feasible.

NLCI accepts typed input only. Future work will use a speech front-end for spoken input. Better coverage of natural language features is also necessary, for instance for handling repetition and parallelism.

3.19 Industry adoption requires empirical evaluations in a real context

Claes Wohlin (Blekinge Institute of Technology – Karlskrona, SE)

License © Creative Commons BY 3.0 Unported license
© Claes Wohlin

Joint work of Wohlin, Claes and Gorschek, Tony

Main reference T. Gorschek, P. Garre, S. Larsson, C. Wohlin, “A Model for Technology Transfer in Practice,” IEEE Software, 23(6):88–95, 2006.

URL <http://dx.doi.org/10.1109/MS.2006.147>

Research solutions are most often prototypes and not industrialized products, and hence research solutions will not be adopted by industry without evaluation in a real context. This requires researchers to 1) have a strategy for evaluation, 2) have a careful methodological approach, 3) take contextual factors into account, and 4) understand representativeness of subjects.

A seven-step evaluation process is proposed: 1) industry-driven research question, 2) study of state-of-the-art and a joint problem formulation with industry, 3) identify a candidate solution, 4) evaluate the solution in an academic setting (risk minimization), 5) conduct a static evaluation of the solution (dry run, i.e. offline from actual development, e.g. with experts from industry), 6) dynamic evaluation for example a case study in an industrial context, and 7) solution is hopefully ready to be released for industry adoption.

To be successful in this endeavor, context has to be taken carefully into account. This includes 1) understand target context (e.g. business vs. open source) and 2) ensure rep-

representative context (e.g. size and type of software system). Furthermore, if assuming use of humans, then subjects in studies are important. This includes two additional factors 3) representativeness of the human subjects, and 4) experience and other relevant attributes of the human subjects.

3.20 Productivity and Data Science for Software Engineering

Thomas Zimmermann (Microsoft Corporation – Redmond, US)

License © Creative Commons BY 3.0 Unported license
© Thomas Zimmermann

In this poster I present my research related to (1) productivity in software engineering and (2) the role of data scientists in software projects.

Participants

- Andrew Begel
Microsoft Res. – Redmond, US
- Alan Blackwell
University of Cambridge, GB
- Margaret M. Burnett
Oregon State University, US
- Rob DeLine
Microsoft Corporation –
Redmond, US
- Yvonne Dittrich
IT Univ. of Copenhagen, DK
- Kathi Fisler
Worcester Polytechnic Inst., US
- Thomas Fritz
Universität Zürich, CH
- Mark Guzdial
Georgia Institute of Technology –
Atlanta, US
- Stefan Hanenberg
Universität Duisburg-Essen, DE
- James D. Herbsleb
Carnegie Mellon University, US
- Johannes Hofmeister
Heidelberg, DE
- Reid Holmes
University of Waterloo, CA
- Christopher D. Hundhausen
Washington State University –
Pullman, US
- Antti-Juhani Kaijanaho
University of Jyväskylä, FI
- Andrew J. Ko
University of Washington –
Seattle, US
- Rainer Koschke
Universität Bremen, DE
- Shriram Krishnamurthi
Brown University –
Providence, US
- Gail C. Murphy
University of British Columbia –
Vancouver, CA
- Emerson Murphy-Hill
North Carolina State University –
Raleigh, US
- Brad A. Myers
Carnegie Mellon University, US
- Barbara Paech
Universität Heidelberg, DE
- Chris Parnin
North Carolina State University –
Raleigh, US
- Lutz Prechelt
FU Berlin, DE
- Peter C. Rigby
Concordia Univ. – Montreal, CA
- Martin Robillard
McGill Univ. – Montreal, CA
- Tobias Röhm
TU München, DE
- Dag Sjøberg
University of Oslo, NO
- Andreas Stefik
Univ. of Nevada – Las Vegas, US
- Harald Störrle
Technical University of Denmark
– Lyngby, DK
- Walter F. Tichy
KIT – Karlsruher Institut für
Technologie, DE
- Claes Wohlin
Blekinge Institute of Technology –
Karlskrona, SE
- Thomas Zimmermann
Microsoft Corporation –
Redmond, US

