DAGSTUHL
REPORTS

Volume 5, Issue 10, October 2015

Dagstuhl Reports, Vol. 5, Issue 10 ISSN 2192-5283

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
http://www.dagstuhl.de/dagpub/2192-5283

Publication date
March, 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`.

Digital Object Identifier: 10.4230/DagRep.5.10.i

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.

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
http://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 15411

Multimodal Manipulation Under Uncertainty

Edited by

Jan Peters[1], Justus Piater[2], Robert Platt[3], and Siddhartha Srinivasa[4]

1 **TU Darmstadt, DE,** mail@jan-peters.net
2 **Universität Innsbruck, AT,** justus.piater@uibk.ac.at
3 **Northeastern University – Boston, US,** rplatt@ccs.neu.edu
4 **Carnegie Mellon University, US,** siddh@cs.cmu.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15411 "Multimodal Manipulation Under Uncertainty". The seminar was organized around brief presentations designed to raise questions and initiate discussions, multiple working groups addressing specific topics, and extensive plenary debates. Section 3 reproduces abstracts of brief presentations, and Section 4 summarizes the results of the working groups.

Seminar October 4–9, 2015 – http://www.dagstuhl.de/15411
1998 ACM Subject Classification I.2.9 Robotics
Keywords and phrases Robotics, Manipulation, Uncertainty, Perception, Computer vision, Range sensing, Tactile sensing
Digital Object Identifier 10.4230/DagRep.5.10.1

1 Executive Summary

Jan Peters
Justus Piater
Robert Platt
Siddhartha Srinivasa

While robots have been used for decades to perform highly specialized tasks in engineered environments, robotic manipulation is still crude and clumsy in settings not specifically designed for robots. There is a huge gap between human and robot capabilities, including actuation, perception, and reasoning. However, recent developments such as low-cost manipulators and sensing technologies place the field in a good position to make progress on robot manipulation in unstructured environments. Various techniques are emerging for computing or inferring grasp configurations based on object identity, shape, or appearance, using simple grippers and robot hands.

Beyond grasping, a key ingredient of sophisticated manipulation is the management of *state information and its uncertainty.* One approach to handling uncertainty is to develop grasping and manipulation skills that are robust to environmental variation. Another approach is to develop methods of interacting with the environment in order to gain task-relevant information, for example, by touching, pushing, changing viewpoint, etc. Managing state information and uncertainty will require a tight combination of perception and planning. When the sensor evidence is unambiguous, the robot needs to be able to recognize that and perform the task accurately and efficiently. When greater uncertainty is present, the robot needs to adjust its actions so that they will succeed in the worst case or it needs to gain additional information in order to improve its situation. Different sensing modalities as well as world models can often be combined to good effect due to their complementary properties.

This seminar discussed research questions and agendas in order to accelerate progress towards robust manipulation under uncertainty, including topics such as the following:

- Is there a master algorithm or are there infinitely many algorithms that solve specialized problems? Can we decompose multimodal manipulation under uncertainty into I/O boxes? If so, what would these be?
- Do we prefer rare-feedback / strong-model or frequent-feedback / weak-model approaches? Is there a sweet spot in between? Is this the way to think about underactuated hands?
- What are useful perceptual representations for manipulation? What should be the relationship between perception and action? What kind of perception is required for reactive systems, planning systems, etc.?
- How do we do deformable-object manipulation? What planning methods, what types of models are appropriate?
- How should we be benchmarking manipulation? What kind of objects; what kind of tasks should be used?
- How should humans and robots collaborate on manipulation tasks? This question includes humans collaborating with autonomous robots as well as partially-autonomous robots acting under human command.

In the area of perception, we concluded that the design of representations remains a central issue. While it would be beneficial to develop representations that encompass multiple levels of abstraction in a coherent fashion, it is also clear that specific visual tasks suggest distinct visual representations.

How useful or limiting is the engineering approach of decomposing functionality into separate modules? Although this question was heavily debated, the majority view among seminar participants was that modules are useful to keep design complexity manageable for humans, and to keep the event horizon manageable for planning systems. It seems that to build more flexible and powerful systems, modules will need to be more strongly interconnected than they typically are these days. Fundamental challenges lie in the specification of each module and of their interconnections. There is a lot of room for creative innovation in this area.

Benchmarking questions were discussed chiefly in the context of the YCB Object Set[1]. Specific benchmarks were suggested and discussed, covering perception and planning in the context of autonomous manipulation.

[1] http://www.ycbbenchmarks.com/

2 Table of Contents

3 Overview of Talks

3.1 Uncertainty during Assistive Manipulation

Brenna Argall (Northwestern University – Evanston, US)

It is a paradox that often the more severe a person's motor impairment, the more challenging it is for them to operate the very assistive machines which might enhance their quality of life. *Assistive manipulators* pose a particular challenge because of their complexity: the dimensionality of the manipulator's control space generally far exceeds the dimensionality of the control signal able to be produced by the human operator (for reasons of motor impairment, or interface limitations). By introducing robotics autonomy and intelligence, we can turn the manipulator into an autonomous robot and offload some of the control burden from the human. Under such an assistance paradigm, *multi-modality* presents itself foremost within the space of *control signals*—since there are multiple (the human, the robotics autonomy) sources controlling the robot platform. *Uncertainty* within the domain of assistive robotic manipulation presents itself in many forms. One way is in the *inference of operator intent*. For the autonomy to provide control assistance requires an idea of the operator's intended task or movement. The aforementioned mismatch in control space complicates and introduces uncertainty into this inference. A second way is in the *estimation of optimal assistance*. Exactly how much assistance is required, or desired, by the operator is a critical unknown. We further expect that the optimal amount of assistance is unique to each human operator, because of the uniqueness of their personal preferences and physical abilities. A third way is in *how to adapt the assistance paradigm*. We expect the optimal assistance solution to change over time, because people's abilities change over time. The right way to adapt the assistance paradigm—autonomously, and without engineer intervention—is unknown, and user-specific. Unknown, yet fundamental—the customization and adaptation of assistance I believe will be critical to the adoption of assistive robots within larger society.

3.2 Learning in Robotics

Leslie Pack Kaelbling (MIT – Cambridge, US)

How can we make robots that are really robust, flexible, and competent in complex relatively unstructured environments? Through a combination of high-level design of algorithms, representations, and structures, on the part of human engineers, and learning and planning, on the part of the robot.

Robots can learn *synthetic* and *analytic* knowledge. Synthetic learning gains actual information about the domain they operate in; it can be represented in terms of policies, values functions, reward models, dynamics models, or observation models; it can be short term (What is the pose of the object in front of me?) or long term (What should the gain on my motor controller be? How does rain affect the arrival time of the people I cook dinner for?); it can be more or less abstract. Analytic learning can be thought of as a kind of compilation or tuning of internal representations: it includes learning to play chess (after you know the rules) or using a slow planner to generate training data for a policy that will be quick to execute.

As domains become more highly variable, more complex, and longer-horizon, I argue that learning structures that can be re-used is critical. Learning a predictive model of kinematics or physics or folk psychology can be re-used over and over with different objectives and can often be adapted with very few training examples. Learning a policy or value function is tied to an objective and will generally be more difficult to transfer, adapt, or re-use. Ultimately, an effective general-purpose robot will need all of these kinds of structures: from fast, specific, low-level policies to slow, abstract, general purpose knowledge and reasoning mechanisms. A critical research question is how to design an architecture that supports these kinds of learning, reasoning, and behavior.

Related questions include:

- What kinds of model representations are most useful for what kinds of problems?
- Can we formulate the objective of a learning problem to include some notion of how the learned structures will be used? (So, for example, it might be useful to learn both a very detailed and a very abstract model for predicting physical interactions of objects, and then be able to employ the most useful one in each circumstance).
- At what point to we really need to address general-purpose reasoning? How should it integrate with the basic planning and learning mechanisms that we use now in robotics?

3.3 Action Selection in Hybrid Spaces

Tomas Lozano-Pérez (MIT – Cambridge, US)

The fundamental planning problems for autonomous manipulation have high-dimensional hybrid state spaces and many actions, also with hybrid parameters. Furthermore, these planning problems, e.g. making dinner, typically have very long planning horizons and take place under substantial uncertainty both in the current state and the result of actions. How do we build effective planners for these problems? We have been exploring approaches built on the following cluster of ideas inspired by AI planning, decision-theoretic planning and motion planning:

- Factored representations of belief space
- Determinize and re-plan for probabilistic planning
- Temporal hierarchies for abstraction
- Implicit representation of pre-images for backchaining
- Relaxed planning problems for heuristic guidance
- Sample-based minimum-constraint removal motion-planning

This talk outlines how these ideas fit together into a coherent whole and discuss strengths and weaknesses.

Related questions include:

- How much of a robot planner could be robot-independent? That is, at what level (if any) could planning effectively become independent of geometry, kinematics, perception, etc.
- How do we effectively combine planning and learning to build autonomous robots?

3.4 Some Basic and/or Old Thoughts on Multimodality and Uncertainty (and New Thoughts on the Amazon Picking Challenge)

Oliver Brock (TU Berlin, DE)

Multimodality

In robotics, *multimodal* is often taken to mean the same as *multi-sensorial.* The question is: what should constitute a modality? If a modality simply is a type of sensor, the matter seems well-understood for a long time. In 1988 Hugh Durrant-Whyte identified three types of multi-sensor integration: competitive, complementary, and cooperative [Sensor Models and Multisensor Integration, IJRR 7(6):97–113, 1988]. But it seems natural to think about the goal of perception when talking about modalities. This view of modalities is what James J. Gibson described in this 1983 book *The Senses Considered as a Perceptual System.* When the senses are viewed a perceptual system, a modality corresponds to a regularity in the sensor data, irrespective of what sensor-type they originated from. This implies that perception should consist of (at least) two layers: one that extracts regularities from multi-sensor streams, and one that leverages these signals appropriately for a particular application. Extracting important regularities, i.e. regularities that are robust and useful for the task of an agent, becomes an important problem in perception. In some way, this changes the general approach: rather than thinking of an application and then identifying the information we might need for it, we should start longing for robust and generally useful regularities in multi-sensorial data. Interestingly, this is exactly what interactive perception is attempting to do.

Uncertainty

There are many approaches to address uncertainty. None of them can be used exclusively to address the uncertainty a real-world robotic system is exposed to. No matter how much effort we invest in modeling uncertainty in a POMDP, there will always be remaining uncertainty that is not reflected in our model. No matter how smart I design my mechanism to suppress uncertainty through clever engineering, there will be situations when the design is insufficient. No matter what assumptions I make about the agent, the world, and their interaction, situations will arise that we not anticipated. At the moment, the key opportunity for addressing uncertainty in real-world robotics is not to advance any one of these possibilities but instead to learn how to cleverly combine them into a robust system.

Amazon Picking Challenge

My lab won the inaugural Amazon Picking Challenge. Of course, the most interesting question is: why? What can be learned from this success? While it is difficult to learn from a single sample, we believe that there were several factors that played a major role.

1. Luck: Back home we tested all five shelf configurations (placement of objects in bins) and the one we had to solve during the competition was the one we performed best on. However, we also performed very well on all the other configurations.
2. Behavior first: Rather than improving components, we always worked on the behavior of the integrated system. We observed that improving the performance of an isolated

component does not necessarily (and maybe rarely) improve the performance of the system containing the component. This has far-reaching consequences. For example, the basic compositional entity (dare I say "module"?) was a behavior of the entire system. Traditionally, these compositional entities are vision systems, planners, controllers, etc.
3. Embodiment: We picked the robot to suit the task. This is reflected in our end-effector (vacuum cleaner with suction cup) and in the fact that we used a mobile base. These are only two choices that ended up making the solutions to other problems much simpler.
4. Prior knowledge: We thought hard about where we can exploit knowledge of the specific problem and the specific setup in our solution. This makes things simpler and therefore more robust, albeit not general, of course.

During the process of preparing for the APC, it became clear to us that "systems papers" in robotics are not very helpful for building systems. Most of them simply describe a particular system, rather than postulating principles of system building that then can be confirmed or disproven by other groups. The robotics community must start writing systems papers to grow a body of knowledge about system building, hopefully leading to some kind of system science.

3.5 On Multifingered Hands and their (Lack of) Industrial Applications

Máximo A. Roa (German Aerospace Center-DLR, DE)

Despite almost 30 years of development in multifingered hands, traditional two-finger parallel jaw grippers are still one of the most common choices for grasping objects in industrial environments. Multifingered hands were developed for solving challenging manipulation tasks, but applications are still marginal, mainly due to the mechanical complexity of the devices, the complex associated control, and their high cost. Hardware is, though, capable of a large variety of interesting behaviors, as demonstrated in teleoperated scenarios of robotic manipulation, or by amputees operating prosthetic devices. The question of the required dexterity in a robotic end effector arises, since simple devices seem to be capable of amazing behaviors, when a human is in control of the actions (planning and execution).

From a robotic perspective, handling uncertainties can be tackled at different levels: hardware, planning or control. From the control point of view, the framework of compliant control applied at object or joint level can cope with deviations of the object pose with respect to the nominal position, leading to robust grasp behaviors. Probabilistic approaches for planning have started to consider uncertainties in the loop. Also, recent hardware advances such as underactuated hands follow the principle of exploiting as far as possible the dynamics of mechanisms to simplify the control tasks, and they favor the exploitation of (instead of avoiding) the contacts with the environment to maximize grasp robustness.

Applications of grasping for manufacturing applications (especially in SMEs) requires still some effort in terms of integration, control, execution and error recovery to guarantee robust applications of robotic technology. The combination of grasp and assembly planning is an example that illustrates the introduction of highly automated workflows for productions of small batches of assembled modular structures. Closing the action-perception loop is crucial for achieving reliability in this domain, using robust error-recovery strategies.

Related questions include:

- How much complexity is required for the end effectors? (Do we really need dexterous hands, or is it enough having task-specific end effectors?)
- How to effectively combine hand and arm dexterity?
- What is the best approach to handle uncertainty in the manipulation process: through hardware, planning, control?
- Is multisensorial perception really needed, or is vision sufficient to successfully perform manipulation tasks?
- Should manipulation problems be solved as combinations of basic skills?

3.6 Haptic Perception and Other Things that Keep Me Up at Night

Veronica J. Santos (University of California – Los Angeles, US)

In the Socratic spirit of this workshop, I will present some of our truths, working assumptions, and challenges in the context of haptic perception for artificial hands. I will provide a brief summary of some (non-intuitive?) insights we have gained from experiments in which a robot hand outfitted with a deformable, multimodal tactile sensor was used to replay human-inspired haptic *exploratory procedures* to perceive edge orientation and fingertip-sized geometric features. I will then give a sneak preview of some new experiments (some more developed than others) on such grand challenge-inspired topics as bimanual manipulation, haptic search within a granular material, and manipulation of deformable materials. These experiments do not yet have "punchlines," but could be used to stimulate discussion of potential pitfalls, alternative approaches, and collaborative extensions of this work. I will end my talk with a list of things that keep me up at night and where I would love to see the field of manipulation in the next 10–20 years.

Truths and working assumptions:

- Perception is an active process.
- Learnable, consistent action-perception relationships are key.
- Solutions can be *bio-inspired* without having to be *biomimetic*.

Related questions include:

- How can we empower robots with a deeper understanding of objects and actions? How can we break away from pre-planned trajectories and teach robots to perceive when a task has been completed or that different actions must be taken to achieve task completion?
- How can we, or should we, teach robots physics? How much physics intuition is needed for a robot to reason about grasp and manipulation tasks?
- What is a practical, useful, (compact? easily searchable? modular?) representation of learned experiences[2] for robotics?
- What machine learning techniques will enable us to achieve online perception and decision-making for interactions with humans at human-like speeds? Or should we not be concerned with speed at this point?

---

[2] Thanks to Leslie Pack Kaelbling, Mehmet Dogar, and Kostas Bekris for this topic.

- How can we discover new solutions for artificial manipulation when current machine learning techniques are limited by our own (often subjective) hand-tuned model parameters, input features, and reward structures?
- How can we extract physical intuition from successful "black box" machine learning approaches?

3.7 Computing Motions for Robots in Healthcare Applications

Ron Alterovitz (University of North Carolina at Chapel Hill, US)

Emerging robots have the potential to improve healthcare delivery, from enabling surgical procedures that are beyond current clinical capabilities to autonomously assisting people with daily tasks in their homes. I will discuss new algorithms to enable medical and assistive robots to safely and semi-autonomously operate inside people's bodies or homes. These algorithms must compensate for uncertainty due to variability in humans and the environment, consider deformations of soft tissues, guarantee safety, and integrate human expertise into the motion planning process.

I will discuss ongoing and future research on motion planning algorithms for new tentacle-like medical instruments, including steerable needles and concentric tube robots, designed for interventional radiology, cardiothoracic surgery, and neurosurgery procedures. These new devices can maneuver around anatomical obstacles to perform procedures at clinical sites inaccessible to traditional straight instruments. To ensure patient safety, our algorithms must explicitly consider uncertainty in motion and sensing to maximize the probability of avoiding obstacles and successfully accomplishing the task. We compute motion policies by integrating sampling-based motion planners, optimal control, and parallel computation. Second, I will discuss new motion planning algorithms for autonomous robotic assistants for helping people with tasks of daily living in the home. I will present demonstration-guided motion planning, an approach in which the robot first learns time-dependent features of an assistive task from human-conducted demonstrations and then autonomously plans motions to accomplish the learned task in modified environments with never-before-seen obstacles.

Related questions include:

- Uncertainty is an inevitable implication of medical robots becoming smaller and gaining degrees of freedom and assistive robots using less precise actuators and sensors to gain compliance and decrease cost. How do we manage uncertainty in robot motion and state estimation in a manner that enables us to provide guarantees on safety?
- Medical robots operate in deformable environments, where both the surrounding soft tissues and the robot itself may deform. How do we model such settings such that we can efficiently compute high-quality motion plans that increase the autonomy and safety of medical robots?
- Physicians don't like to hand over all control to an autonomous agent, but like receiving assistance that makes their job easier. How should robots and physicians effectively share autonomy during surgery?

3.8 Data, Data, Data – How much of it do we really need in robotics?

Jeannette Bohg (MPI für Intelligente Systeme – Tübingen, DE)

Large-scale, labeled databases for classic tasks in Computer Vision or speech recognition and the ability to learn from them provide the key to the state-of-the-art performance on many benchmarks that are important in these areas. In robotics however, databases of this magnitude are rare which may be due to the effort that would be required to collect them. Traditionally, robotics is much more driven by models that are for example based on first principles. Often these provide very useful abstractions of dynamical systems to develop robust controllers. But in many cases, we have seen that the underlying assumptions of such a models do not hold in the real world. This has for example been the case for grasping. In some of our recent work, we assembled a large-scale database for learning to grasp given only partial and noisy data of the object. The data points are automatically annotated using physics simulation of the grasp. This type of data helps to synthesize stable grasps and even to predict some of the latent object properties like global object shape, category or contact locations.

However, I wonder about two related questions:

1. What is the right mixture of first-principle modelling and learning? Do we want to learn entirely from scratch to not bias the resulting model with our potentially too restrictive ideas? Or do we want to learn isolated parameters of an otherwise fixed model? In what situations is learning in one way or the other preferred?
2. Related to the above, how much data do we really need? How should it be collected: incrementally or in batch? What is the important data to collect: all the modalities or some intuitively important ones? Should it be time-series data or data collected at a particular point in time?

The relation between these two questions lies in the apparent trade-off between including expert knowledge in our models and the amount of data required to learn the remaining open parameters.

4 Working Groups

4.1 Perception

- What are useful perceptual representations for manipulation?
- What should be the relationship between perception and action?
- What kind of perception is required for reactive systems, planning systems, etc.?

4.1.1 Group 1

We began by discussing the uses of perception: disambiguation, association, and creation of a unified world model for decision-making. We discussed how perception representation must include uncertainty, and that tactile and/or proximity sensing could be used to reduce uncertainty. The value of tactile sensing seems to be (unnecessarily) dependent upon the approach to manipulation. For example, reliance on contact location information limits the usability of certain tactile sensor technologies that might otherwise be useful if a different approach were taken. There was disagreement on the richness of tactile information needed for manipulation. We discussed whether complete 3D models of objects are necessary to grasp and manipulate the environment. Alternative approaches were suggested, such as focusing on labeling of objects, or starting with a partial model (e.g. incomplete surface models) and using tactile sensing to adjust grasp after contact is made. We discussed reactive systems as enacting direct mappings of inputs to outputs that do not project forward in time. However, planning systems can be implemented as reactive systems (e.g. reinforcement learning).

Different approaches to perception were discussed. One approach is to broadly search for patterns in sensor data. Another approach is to look specifically for relevant task-based information. We discussed the degree to which sensor data must be processed for use in the world model. Classical AI approaches use a *lazy* approach to perception in which data are stored and accessed when needed. In contrast, control theory approaches use a *greedy* approach in which filters are used to maintain online, on-demand, updated models of the world at all times. We discussed whether awareness was necessary for perception. Is it possible to have physical, but not computational awareness? For example, a jamming gripper may not be aware of its shape, and yet it still reacts and changes shape in response to stimuli.

We concluded that the explicit representation of uncertainty is essential for combining representations of the world. The network for fusing different representations of the world and queries of the environment can be task-specific. Action and perception should be considered jointly. There is no "correct" approach to perception, as the approach depends on the application.

4.1.2 Group 2

We observed that object-action representations span multiple levels of abstraction. Depending on the information available to the robot and the task at hand, the link between perception and action can be direct (e.g., motor reflex linked to a specific tactile pattern), or it can pass through abstracted representations, for instance a 3D vector representation of a world contact, surface patches with surface normals (used for hand-object interaction models, as in the work of Kopicki and Wyatt), complete object surface representations (as in grasp densities), symbolic labeling, generic hand postures or prototypical motor patterns.

The outcome of our discussion is that it would be profitable to develop representations that encompass multiple levels of abstractions, instead of using distinct representations for different levels. The representation should also allow for long term information gathering (graceful integration), and lend itself to choosing actions that disambiguate perception.

4.1.3 Group 3

The perceptual representation to use depends on the task at hand. Even during a simple grasping operation, we can create a matrix mapping each subtask to a perceptual representation: *a geometric representation* during motion planning, *an appearance-based representation* during visual servoing to align the hand with the object, *a direct representation* during the guarded move to make the first contact with the object, and so on. Some of the important and useful representations include: spatial probabilistic representations over poses, spatial probabilistic representations over occupancy grids, symbolic representations, affordances, appearance/material/inertial representations, worst/best-case geometry representations, direct representations ("world is its own model"), and predictive state representations.

4.2 General vs. Specific Solutions

4.2.1 Group A

- Is there a master algorithm or are there infinitely many algorithms that solve specialized problems?
- Can we decompose multimodal perception under uncertainty into a simple system of separate, interconnected I/O modules? If so, what would these be?

Assuming that a master algorithm composed of isolated submodules exists, one can start thinking of how such an algorithm could look like. This master algorithm would consist of separate modules for computer vision, control and others. For such a modular structure, each module is working independently of the other modules to some degree (e.g. a module can simply be replaced without any effort). However, the question came up, how modularity is defined in this setting. We define a module by the number of connections inside and outside the module (e.g. a component is modular, if some subparts have high inter-connectivity but the number of connections to other components is low). However, it still remains unclear, if such a definition is meaningful as there are different levels of modularity. Almost every system can be formulated in a modular way, even though it isn't (e.g. the human brain is known not to be strictly modular; however, on a lower level it consists of many neurons that can be seen as modules). In order to decide for the likelihood of a strictly modular approach to be successful we looked at the human brain. The brain is clearly not modular in a strict sense. Many synapses are connecting neurons from different areas within the brain. Still, one can define several areas that mainly seem to be responsible for a specific purpose. Therefore, we do not reject the idea of modularity (and the currently dominant engineering approach in robotics), but we propose the significant extension of having higher inter-connection between the modules. Each module incorporates a different prior on the input data and produces a prediction based upon it. In order to obtain much more robust functionality, these outputs could be fed back to the other modules. Such a paradigm should be able to cancel out noise significantly. In the perception domain, this relates to the fusion of different sensing modalities. For humans it is known that these affect each other. For example a human can smell, see, feel and even taste fire, where each of these separate modalities can be used to predict much more precisely if fire is present or not.

4.2.2 Group B

Can we decompose multimodal manipulation under uncertainty into I/O boxes? – Our conclusion was: yes. Large successful engineered systems typically can be represented as boxes. A key challenge is how to create them and connect them.

Creating *useful* boxes is challenging. A box should be defined by a task or a well-defined problem; it should have preconditions and postconditions (i.e., a goal) that are well defined and we should be able to develop an algorithm for it. A box needs to be thought about based on how it interacts with other boxes; if one decides on a box without thinking about other boxes, one cannot define a useful box. Smaller boxes are better, e.g., ape's primitive motions, a learned subtask, and a push grasp.

When creating boxes, a consideration is whether to take a top-down vs. bottom up approach. In the top-down approach, one starts with a general approach (e.g., a hierarchy) and then creates boxes for various operations in a hierarchy. In the bottom-up approach, one builds boxes for certain skills, then composes them for more complex tasks.

Deciding what boxes are needed is challenging. A decomposition of a task into finite boxes restricts the possible choices that a system has. By using a decomposition, the gain is simplification, and the loss is that we restrict the system from doing certain things. For example, for performing tasks in clutter, if we only have boxes for push grasps and some simple transitions, then we eliminate the possibility of doing something entirely different that might be useful, such as throwing an object. Also, boxes may be myopic, making it difficult to do error attribution and recovery.

4.2.3 Group C

Can we decompose multimodal manipulation under uncertainty into I/O boxes? If so, what would these be?

There are results from neuroscience indicating a strong modularity in the human sensorimotor areas. E.g. the neural signals related to individual fingers are independent. I.e., these signals look the same if a person hold a cup with her own e.g. 2 fingers vs. 2 persons with a finger each. So there is some evidence from the biological side for modularity.

The next question then is, should we think in terms of boxes (and their connections) or always consider loops including feedback signals as the basic building block? The latter would seem a bit like the subsumption architecture, which worked fine for simple, reflex behaviour based agents, but fell short for more complex behaviour requiring a larger planning horizon. We do need hierarchies, with levels of abstraction to keep the planning horizon small. One problem with such is that they can be too restrictive. The designer of such a hierarchy might have been thinking about a certain set of problems, for which the particular hierarchy seemed a "natural" way to organise the various modules/boxes. But this might prevent the system from solving slightly different problems. Is there even a necessary, "natural" modularity? Different types (regarding time spans, horizons) of tasks require different representations and thus *levels*. Differences are 2D/3D, work space/task space, short/long horizon, fast/slow.

So as long as there is no one, obvious architecture/hierarchy/nodularity we will have several competing propositions. How can we make and measure methodological progress? I.e., how do we know whether a certain architecture is better than another? This is very difficult to ascertain. Also because one can not simply connect boxes in a different way to obtain a different architecture. There is an inherent connection between boxes and architectures. A decision for a specific (type of) box is already a decision about the architecture in which the box is going to live.

One pragmatic option is to take stock of what functionalities/modules/boxes are available and selecting a path connecting all the modules necessary to solve a concrete problem. What follows is an exemplary (of course incomplete) list of modules:

- Segmentation: in: RGBD; out: segments
- Recognition: RGBD, model; out: instance, pose
- Classification: in RGBD, model, priors (segments); out: classified objects
- Tracking: in: RGBD, model, prev. pose; out: next pose
- Reinforcement learning: any observation, hypothesis class; out: actions
- Supervised learning: in: D, y; out: y
- Unsupervised learning: in: D; out: f(D)
- Object state filter: in: seq. of obj detections; out: set of obj. hyp.
- Metric SLAM: in: RGBD, laser; out: occupancy grid
- Motion planning: model of world, model of robot; out: path
- Trajectory Controller: in: goal x; out: control signal
- Grasp generator R: in: model, RGBD, target object category; out: hand pose
- Grasp generator M: in: model, RGBD, grasp category; out: trajectory

- Grasp generator J: in: model, RGBD; out: hand pose
- Grasp controller: in: joint angles, torques, tactile readings; out: joint torques
- Non-prehensile single finger manipulation: in: tactile readings, joint torques; out: joint velocities

For all these modules it is critical to make all implicit assumptions which the designer made very explicit. Otherwise, connecting a set of modules will lead to a very brittle system, failing for no obvious reason as soon as any assumption is violated.

Is there a master algorithm or are there infinitely many algorithms that solve specialized problems?

So having subscribed to the idea that there are many boxes, connected within this or that architecture to form a complete system, the question is: Do we have many similar boxes (one for grasping door handles, one for grasping mugs, one for grasping cloth, ...), or is each box very general (a master type algorithm)? In the former case one would take whatever boxes available and use a good architecture to connect them smartly. In the latter case one would first aim for universally usable boxes, that always work (independent of task), and then build an architecture around these.

Not that we do not necessarily have to aim for one single algorithm, but hopefully we can formulate one framework (general unified strategy), which can then be specialized for specific tasks. Note that the SLAM community seems to have converged to one such well-understood framework. Can we hope for something similar regarding manipulation?

4.2.4 Group D

Is there a master algorithm or are there infinitely many algorithms that solve specialized problems?

There is not enough evidence to answer this question. While human studies would seem to suggest that a master algorithm can exist, the conclusion is far from irrefutable, and one can easily veer into "pop psychology" when trying to answer the question. More clear seems to be the fact that, so far, the robotics field has had success by using many algorithms, loosely following the "boxes" paradigm. Examples are too numerous to build complete lists.

Can we decompose multimodal manipulation under uncertainty into I/O boxes? If so, what would these be?

As mentioned above, this approach has served the field well so far. Still, many caveats must be mentioned. It is important not to confuse computational modularity with task-space modularity (in other words, it's one thing for the APIs of two modules to match, but do the assumptions they make about the environment also match?). Furthermore, independent design of modules that are meant to be combined later is a difficult undertaking. One must remember that it is possible to achieve a bad design when using good modules. It is also important to remember that interconnections matter. For example, it is a topic of discussion whether perception and action can be separated into two boxes. However, even if the answer is affirmative, the connection between such boxes can not be a single point in state space (i.e. perception uniquely identifies the state of the world, then passes it on to action)!

4.3 Protocols and Benchmarks

4.3.1 Perception

We began by identifying existing benchmarking and challenges in the computer vision community: ICCV Workshop on Recovering 6D Object Pose, and the Visual Object Tracking Challenge. We then discussed similar efforts in the robotics community: YCB Object Set, 3D object databases, and the Amazon Picking Challenge. The sharing of datasets was briefly discussed. While camera images are easily shared and pooled, it was less clear how tactile datasets could be shared. Should action information, such as force and motion time histories, be provided along with tactile data?

We discussed the feasibility of experiments conducted over the internet. It was recommended that the robotics community appeal to federal institutions (e.g. NIST) and/or companies (e.g. KUKA, ABB) to house, host, and maintain hardware for communal benchmarking use. It was unclear which hardware would be appropriate and whether the effects of the hardware could be eliminated. One idea was to initially have two pathways and see which (or both) gain traction with the robotics community. Pathway 1 would use a single set of universal hardware for communal use. Pathway 2 would use a single task and no constraints on hardware selection. If some groups wanted to focus on one aspect of a manipulation experiments could be designed based on assumed modules of capability. For example, a specific grasp planner would be provided and researchers would test their ability to provide accurate inputs (e.g. 6D object pose) to the grasp planner module. We discussed how performance could be evaluated. Points were deemed to be too arbitrary. The time to completion was identified one objective measurement. Statistics could be used to report accuracy and frequency of success. It was unclear when and how often to assess performance during a single task.

We discussed the possible outputs of a perception system. For instance, should the perception system output 3D coordinates, a grasp pre-shape, a delta change in movement, etc.? The exact output would likely depend on the hardware, approach, and whether the system is open- or closed-loop. We discussed how to benchmark tactile sensing-related tasks, and whether it would be premature to do so. Care should be taken to prevent the benchmarking effort from turning into a test of the sensor technology itself. A benchmarking protocol would need to focus on perceptual information and not raw sensor data. Protocols could be designed that purposely occlude or disallow computer vision. Protocols could test the ability to perform tasks that require detection of discrete events (e.g. slip), and/or the ability to performance tasks that require tracking of forces. Tactile object recognition was one suggested task.

We identified two candidate benchmarking tasks:

1. Picking up an object from a scene and dropping in a bin
   - Input: Object to be selected, Point cloud
   - Output: 6D object pose in space, Gripper pose with respect to object
   - Constraints: fixed hardware, fixed planning, fixed grasping algorithm
   - Evaluation metric: Grasp and pick success
2. Placing object
   - Input: Point cloud
   - Output: 6D object pose in space, Gripper pose with respect to object
   - Evaluation metric: Placement accuracy, success

4.3.2 Planning

Planning benchmarks can defined in terms of complexity along the following axes:

- the type of *robot*: one arm vs. multiple arms,
- the type of *interaction*: basic pick-and-place vs. dextrous manipulation,
- the types of *objects* being manipulated: rigid, articulated, deformable, or a combination,
- the type of *uncertainty* (in action and perception): none, corrected/bounded by special moves/controllers, explicit planning in belief space.

Orthogonal to a categorization of planning benchmarks along axes of complexity, we can also define a taxonomy of robot planning capabilities. Planning algorithms will typically focus only on a subset of such capabilities. Examples of such capabilities include:

- collision-free point to point planning,
- task planning, assembly, and object rearrangement,
- grasping,
- bimanual manipulation: coordinated motion (closed kinematic chains), concurrent motion,
- manipulating deformable objects and articulated objects,
- planning in dynamic (i.e., time-varying) scenes,
- planning dynamic motions (e.g., throwing, waiter-like motion for carrying a tray),
- handover of objects (to a human),
- object search (necessarily includes sensing),
- planning for tasks that require tool use.

Considerations for a useful benchmark include that it should be predictive of how well a planning problem would work on other, similar problems. The benchmark should be well-defined so that scores can be compared in a meaningful way across contributors. On the other hand, it should be made difficult for a hacky, special-purpose planner to do well. One way to do this is introduce some randomization in the benchmark's initial conditions and report average score across a number of trials. Simulation can also be a useful to evaluate parameter / pose sensitivity of planner for a given benchmark.

We arrived at the following three benchmark concepts:

1. **Page turning.** This would test the ability to manipulate deformable objects. Performance can be measured in terms of number of pages turned within some time.
2. **Object search on a shelf.** Here we can define several variants. For example, the initial object arrangements could be well-defined, but not be given to the robot. This means that the robot needs to have some exploration strategy. In other words, it would also test perception. In simulation, perfect perception can be faked to measure just the effectiveness of the planner. Alternatively, the exact location of each object can be specified and the task would "simply" be to remove one of the objects, which would entail having to move other objects out of the way. By placing the objects in a cluttered environment (e.g., by placing objects within a small cubby), the problem can be made very hard.
3. **Assembly of Duplo bricks.** This would benchmark task planning, pick-and-place planning, compliant motion, and force control in mating the Duplo pieces. The problem can be made even harder by requiring bimanual in-hand assembly (i.e., using one hand to pick up pieces while the other hand holds the partial assembly). The problem can be made slightly easier by specifying the task plan. The initial placement of the Duplo pieces can be random (to avoid hardcoding the moves) or explicitly defined. The desired assemblies can be divided into easy/medium/hard categories. The initial arrangement blocks can be made easy (all blocks right side up) or hard (blocks sideways or upside down).

5 Open Problems

The group identified several open problems:

1. **Benchmarks for robotic manipulation:** A key challenge is in developing benchmark objects and protocols that a) are of interest to all communities (planning, perception, control, learning), b) generalizable across robot platforms, and c) accessible so that every group can try them out.
2. **Developing building blocks:** Manipulation requires the tight integration of several components. A key challenge is in developing an open-source architecture that enables several groups to contribute specific modules they are expert in, and harness modules written by other groups.
3. **Perception for manipulation:** The perception community has several excellent benchmarks for computer vision. However, manipulation has specific demands: dealing with clutter, and outputting the 6D pose of objects, among others. Many manipulation systems suffer from poor perception and an open challenge is in harnessing the perception community to address the specific demands of manipulation.

6 Panel Discussions

Our seminar had several interesting discussions (detailed above) in smaller groups. We found this to be more useful because a) it allowed everyone to have a greater opportunity to speak (as our total group size was large), and b) it allowed much more focussed discussions and deep dives on a particular topic.

Participants

- Ron Alterovitz
University of North Carolina at Chapel Hill, US
- Brenna D. Argall
Northwestern University – Evanston, US
- Yasemin Bekiroglu
KTH Royal Institute of Technology – Stockholm, SE
- Kostas Bekris
Rutgers Univ. – Piscataway, US
- Dmitry Berenson
Worcester Polytechnic Inst., US
- Bastian Bischoff
Robert Bosch GmbH – Stuttgart, DE
- Jeannette Bohg
MPI für Intelligente Systeme – Tübingen, DE
- Oliver Brock
TU Berlin, DE
- Matei Ciocarlie
Columbia University, US
- Fan Dai
ABB Corporate Research, DE
- Renaud Detry
University of Liège, BE
- Mehmet R. Dogar
University of Leeds, GB
- Aaron M. Dollar
Yale University, US
- Roderic A. Grupen
University of Massachusetts – Amherst, US
- Simon Hangl
Universität Innsbruck, AT
- David Hsu
National Univ. of Singapore, SG
- Leslie Pack Kaelbling
MIT – Cambridge, US
- Marek S. Kopicki
University of Birmingham, GB
- Dirk Kraft
University of Southern Denmark – Odense, DK
- Norbert Krüger
University of Southern Denmark – Odense, DK
- Ville Kyrki
Aalto University, FI
- Ales Leonardis
University of Birmingham, GB
- Shuai Li
Rensselaer Polytechnic, US
- Maxim Likhachev
Carnegie Mellon University, US
- Tomás Lozano-Pérez
MIT – Cambridge, US
- Matthew T. Mason
Carnegie Mellon University – Pittsburgh, US
- Mark Moll
Rice University – Houston, US
- Duy Nguyen-Tuong
Robert Bosch GmbH – Schwieberdingen, DE
- Erhan Öztop
Özyegin Univ. – Istanbul, TR
- Jan Peters
TU Darmstadt, DE
- Justus Piater
Universität Innsbruck, AT
- Robert Platt
Northeastern University – Boston, US
- Maximo A. Roa
German Aerospace Center-DLR, DE
- Veronica Santos
University of California – Los Angeles, US
- Siddhartha Srinivasa
Carnegie Mellon University, US
- Ales Ude
Jozef Stefan Institute – Ljubljana, SI
- Francisco Valero Cuevas
USC – Los Angeles, US
- Jeremy L. Wyatt
University of Birmingham, GB
- Michael Zillich
TU Wien, AT

Report from Dagstuhl Seminar 15412

Dynamic Traffic Models in Transportation Science

Edited by

José R. Correa[1], Tobias Harks[2], Kai Nagel[3], Britta Peis[4], and Martin Skutella[5]

1 **University of Chile – Santiago de Chile, CL,** correa@uchile.cl
2 **Maastricht University, NL,** t.harks@maastrichtuniversity.nl
3 **TU Berlin, DE,** nagel@vsp.tu-berlin.de
4 **RWTH Aachen, DE,** britta.peis@oms.rwth-aachen.de
5 **TU Berlin, DE,** martin.skutella@tu-berlin.de

Abstract

Traffic assignment models are crucial for traffic planners to be able to predict traffic distributions, especially, in light of possible changes of the infrastructure, e.g., road constructions, traffic light controls, etc. The starting point of the seminar was the observation that there is a trend in the transportation community (science as well as industry) to base such predictions on complex computer-based simulations that are capable of resolving many elements of a real transportation system. On the other hand, within the past few years, the theory of dynamic traffic assignments in terms of equilibrium existence and equilibrium computation has not matured to the point matching the model complexity inherent in simulations. In view of the above, this interdisciplinary seminar brought together leading scientists in the areas *traffic simulations*, *algorithmic game theory* and *dynamic traffic assignment* as well as people from industry with strong scientific background who identified possible ways to bridge the described gap.

Seminar October 4–9, 2015 – http://www.dagstuhl.de/15412
1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity, I.6 Simulation and Modelling, E.1 Data Structures
Keywords and phrases Dynamic traffic equilibria, Complexity of equilibrium computation, Simulation, Dynamic network flow theory, Network optimization
Digital Object Identifier 10.4230/DagRep.5.10.19
Edited in cooperation with Miriam Schlöter

1 Executive Summary

José Correa
Tobias Harks
Kai Nagel
Britta Peis
Martin Skutella

Traffic assignment models play an important role for traffic planers to predict traffic distributions, especially, in light of possible changes of the infrastructure, e.g., road constructions, traffic light controls, etc. The prevailing *mathematical* approaches used in the transportation science literature to predict such distributions can be roughly classified into static traffic assignment models based on aggregated static multi-commodity flow formulations and dynamic traffic assignment (DTA) models based on the methodology of flows over time. While static models have seen several decades of development and practical use, they abstract away too many important details and, thus, become less attractive. On the other hand, dynamic models are known to be notoriously hard to analyze in terms of existence, uniqueness and computability of dynamic equilibria.

In light of the prevailing computational difficulties for realistic-sized networks, the systematic optimization of such networks (e.g., by designing the network infrastructure, link tolls, or traffic light controls) becomes even more challenging as the resulting mathematical programs with equilibrium constraints contain already in the lower level presumably “hard” optimization-, complementarity- or variational inequality problems; not to speak of the resulting optimization problem for the first level.

On the other hand, there is a trend in the transportation science community to use *large-scale computer-based microsimulations* for predicting traffic distributions. The striking advantage of microscopic simulations over DTA models is that the latter usually ignore the feedback of changing network conditions on user behavior dimensions such as flexible departure time choice, mode choice, activity schedule choice, and such. Current simulation tools integrate all these dimensions and many more. The increasing model complexity, however, is by far not matched by the existing theory of dynamic traffic assignments. Against this background, the seminar provided (partial) answers to questions of the following type:

- Under which conditions do microscopic simulation models and dynamic traffic assignment models admit an equilibrium?
- Is an equilibrium efficiently (polynomial time) computable?
- Which models lead to multiple equilibria and how do the parameters of a learning process influence the resulting equilibrium outcome?
- What are the implications of possible intractability results (PPAD-hardness) on the plausibility of existing models?
- how do we compute optimal (or approximatively) network designs or traffic light controls subject to dynamic equilibrium constraints in polynomial time?

The seminar brought together leading researchers from three different communities – Simulations (SIM), Dynamic Traffic Assignment (DTA) and Algorithmic Game Theory (AGT) – and identified ways to narrow the existing gap between complex simulation based models and the existing theory. Among other points, the seminar initiated a systematic study of the complexity of equilibrium computations for DTA models – which is the core task when resolving dynamic traffic assignment problems. Equilibrium computation and its complexity status is a core topic in AGT. The seminar provided an excellent forum for a discourse of these questions between the DTA, SIM and AGT community which initiated several novel research questions and directions. The seminar also stimulated a conceptual discourse regarding the validity of DTA and microscopic simulation models in terms of their predictive power and use for optimization based approaches.

Overall, the seminar was a big success both in terms of stimulating new and very fruitful collaborations between so far separate communities and also with respect to novel insights and results on traffic equilibria and related concepts. We got enthusiastic feedback from many participants which is also reflected in the survey conducted by Dagstuhl.

2 Table of Contents

3 Overview of Talks

3.1 A Stackelberg Strategy for Routing Flow over Time

Umang Bhaskar (TIFR Mumbai, IN)

Joint work of Bhaskar, Umang; Fleischer, Lisa; Anshelevich, Elliot

We study the efficiency of routing games in a dynamic queuing model introduced by Koch and Skutella. Prior work on routing games with static flows assumes that users care about either their maximum delay or their total delay. Both these measures are surrogates for measuring how long it takes to get all of a user's traffic through the network, and the use of the dynamic queuing model allows us to directly address this objective. We show that in this model, by reducing network capacity judiciously, the network owner can ensure that the equilibrium is no worse than a small constant times the optimal in the original network, for two natural measures of optimality. These are the first upper bounds on the price of anarchy in this model for general networks.

3.2 Dynamic Equilibrium in Network Flows (a limited survey)

Roberto Cominetti (University of Chile – Santiago de Chile, CL)

We present an overview of the alternative equilibrium models that have been proposed for dynamic flows: volume-delay functions, LWR link dynamics, fluid queues, abstract models. We discuss both path-based and link-based models and revise the known existence results, pointing out some limitations as well as some open issues. In the second part of the talk we will focus on the computation of equilibria for the fluid queue model, raising a number of basic questions that remain unanswered.

References

1 Roberto Cominetti, José Correa, Omar Larré. *Dynamic equilibria in fluid queuing networks*. Operations Research 63(1):21–34, 2015.

3.3 Tight Bounds for Cost-Sharing in Weighted Congestion Games

Martin Gairing (University of Liverpool, GB)

Joint work of Gairing, Martin; Kollias, Konstantinos ; Kotsialou, Grammateia
Main reference M. Gairing, K. Kollias, G. Kotsialou, "Tight Bounds for Cost-Sharing in Weighted Congestion Games," in Proc. of the 42nd Int'l Colloquium on Automata, Languages, and Programming (ICALP'15),, LNCS, Vol. 9135, pp. 626–637, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-662-47666-6_50

This work studies the price of anarchy and the price of stability of cost-sharing methods in weighted congestion games. We require that our cost-sharing method and our set of cost functions satisfy certain natural conditions and we present general tight price of anarchy bounds, which are robust and apply to general equilibrium concepts. We then turn to the price of stability and prove an upper bound for the Shapley value cost-sharing method, which holds for general sets of cost functions and which is tight in special cases of interest, such as bounded degree polynomials. Also for bounded degree polynomials, we close this talk with a somehow surprising result, showing that a slight deviation from the Shapley value has a huge impact on the price of stability. In fact, for this case, the price of stability becomes as bad as the price of anarchy.

3.4 Cascading to Equilibrium: Hydraulic Computation of Equilibria in Resource Selection Games

Yannai A Gonczarowski (The Hebrew University of Jerusalem, IL)

Joint work of Gonczarowski, Yannai A.; Tennenholtz, Moshe
Main reference Y. A. Gonczarowski, M. Tennenholtz, "Cascading to Equilibrium: Hydraulic Computation of Equilibria in Resource Selection Games," arXiv:1403.7605v6 [cs.GT], 2015.
URL http://arxiv.org/abs/1403.7605v6

Drawing intuition from a (physical) hydraulic system, we present a novel framework, constructively showing the existence of a strong Nash equilibrium in resource selection games (i.e., asymmetric singleton congestion games) with nonatomic players, the coincidence of strong equilibria and Nash equilibria in such games, and the uniqueness of the cost of each given resource across all Nash equilibria.

Our proofs allow for explicit calculation of Nash equilibria and for explicit and direct calculation of the resulting (unique) costs of resources, and do not hinge on any fixed-point theorem, on the Minimax theorem or any equivalent result, on linear programming, or on the existence of a potential (though our analysis does provide powerful insights into the potential, via a natural concrete physical interpretation).

A generalization of resource selection games, called resource selection games with I.D.-dependent weighting, is defined, and the results are extended to this family, showing the existence of strong equilibria, and showing that while resource costs are no longer unique across Nash equilibria in games of this family, they are nonetheless unique across all strong Nash equilibria, drawing a novel fundamental connection between group deviation and I.D.-congestion. A natural application of the resulting machinery to a large class of constraint-satisfaction problems is also described.

3.5 Coordination Mechanisms and Routing over Time

Martin Hoefer (Universität des Saarlandes, DE)

Joint work of Hoefer, Martin; Mirrokni, Vahab; Röglin, Heiko; Teng, Shang-Hua
Main reference M. Hoefer, V. Mirrokni, H. Röglin, S.-H. Teng, "Competitive Routing over Time," Theoretical Computer Science, 412(39):5420–5432, 2011.
URL http://dx.doi.org/10.1016/j.tcs.2011.05.055

We consider a framework that enhances network congestion games with a notion of time. Our temporal network congestion games are based on coordination mechanisms – local policies that allow to sequentialize traffic on the edges. We study existence and complexity properties of pure Nash equilibria and best-response strategies for linear latency functions. In some cases our results can be used to characterize convergence properties of various improvement dynamics, by which the population of players can reach equilibrium in a distributed fashion.

3.6 Complexity and Approximation of the Continuous Network Design Problem

Max Klimm (TU Berlin, DE)

Joint work of Gairing, Martin; Harks, Tobias; Klimm, Max
Main reference M. Gairing, T. Harks, M. Klimm, "Complexity and Approximation of the Continuous Network Design Problem," in Approximation, Randomization, and Combinatorial Optimization – Algorithms and Techniques (APPROX/RANDOM'14) – Proc. of the 17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems and 18th Int'l Workshop on Randomization and Computation, LIPIcs, Vol. 28, pp. 226–241, Schloss Dagstuhl, 2014.
URL http://dx.doi.org/10.4230/LIPIcs.APPROX-RANDOM.2014.226

We revisit the classical (bilevel) continuous network design problem. Given a graph for which the latency of each edge depends on the ratio of the edge flow and the capacity installed, the goal is to find an optimal investment in edge capacities so as to minimize the sum of the routing costs of the induced Wardrop equilibrium and the investment costs for installing the edge's capacities. We show that continuous network design is APX-hard. As for the approximation of the problem, we provide a detailed analysis for a heuristic studied by Marcotte [1]. Then, we propose a different algorithm and prove that using the better of the two algorithms results in improved approximation guarantees.

References

1 P. Marcotte. *Network design problem with congestion effects: A case of bilevel programming*. Math. Progam., 34:142–162, 1986.

3.7 Present and Future of Dynamic Traffic at TomTom

Felix Koenig (TomTom – Berlin, DE)

TomTom is a world market leader in both dynamic routing and real-time traffic information. Independent research proves that in the present, drivers with different quality of traffic information, and different levels of sophistication in dynamic routing, experience different travel times between common origin and destination. This seems to contradict the common assumption that drivers in dense traffic tend to form a Nash equilibrium.
In future scenarios for transport in Smart Cities, mobility providers might control the routes taken by a significant share of vehicles on the road. This highlights the importance of coalition games in the future of dynamic traffic and can open up new research opportunities.

3.8 Detecting Braess's paradox in MATSim

Kai Nagel (TU Berlin, DE)

Joint work of Thunig, Theresa; Nagel, Kai

MATSim (Multi-agent transport simulation) has, as network loading model, a relatively simple queue model, against which agents learn until they are roughly in a Nash Equilibrium. Moving the well-known Braess example into that simulation leads to some maybe unexpected consequences including the fact that the transient is rather different from the steady state, and the Braess paradoxon (that the additional road REDUCES overall system capacity) may not show up or not show up fully. Clearly, all of this can be explained and understood, but it points to the fact that seemingly small details in the representation of the traffic dynamics may lead to rather different solutions.

3.9 Congestion games with strategic departures

Marco Scarsini (LUISS Guido Carli – Rome, IT)

We consider a discrete-time atomic congestion game with a single source-destination pair and a finite number of players who have to reach their destination by an exogenously fixed time t_0 and can decide at what time to enter the system in order to achieve their goal. Players who arrive late incur a huge cost. Each edge has a capacity and a delay.

We start considering a very simple network with just one edge and we prove that the game has no pure Nash equilibrium. The worst mixed Nash equilibrium is symmetric and socially quite bad, since each player pays her minmax cost. The best Nash equilibrium is only marginally better. The social optimum is achieved by spreading players at capacity over time. The price of Anarchy is approximately 2.

There exists a correlated equilibrium whose cost is quite close to the optimum. Therefore a planner could (almost) achieve efficiency in equilibrium by providing a suitable correlation device.

We plan to extend the results to more general networks.

3.10 Computing Earliest Arrival Flows in Polynomial Space

Miriam Schlöter (TU Berlin, DE)

In the last years, the number of evacuations for example in areas endangered by natural disasters has increased. Thus, a better evacuation planning before the emergency occurs is of great interest. The core of evacuation planning is captured by earliest arrival flows. Given a network $\mathcal{N}$ with capacities and transit times on the arcs, a subset of source nodes with supplies and a single sink node, an earliest arrival flow is a dynamic flow in $\mathcal{N}$ such that the total amount of flow that has arrived at the sink is maximal for all points in time.

In networks with a single source earliest arrival flows can be computed in polynomial space using the successive shortest path algorithm.

For networks with multiple sources Nadine Baumann and Martin Skutella developed an algorithm to compute earliest arrival flows [1]. Their algorithm consists of two parts: At first the earliest arrival pattern is computed and after that using the breakpoints of the pattern the actual earliest arrival flow is derived.

While the first part of the algorithm only works on the original network, the second part requires attaching an additional sink to the network for every breakpoint of the earliest arrival pattern. As the earliest arrival pattern in the worst case can have exponentially many breakpoints, the resulting network also can get exponentially large.

During the computation of the pattern the times at which the sources run empty in an earliest arrival flow are computed. Mainly making use of these times, we present an algorithm to compute earliest arrival flows in networks with multiple sources which only requires polynomial expansion of the original network.

References

1 N. Baumann. *Solving evacuations problems efficiently – earliest arrival flows with multiple sources*. In FOCS Proceedings, 399–408, 2006.

3.11 Atomic Selfish Routing over Time

Daniel Schmand (RWTH Aachen, DE)

In this talk we discuss a model for atomic selfish routing games over time. We assume that we are given some demand which travels unsplittable and selfishly from a given origin to a destination over time. In addition to that we are given capacities on the directed edges, which are an upper bound on the amount of flow that is allowed to enter an edge at a certain time. This may force some flow to wait on intermediate nodes of the route.

In the first part of the talk we try to discuss the model and the question how to design good scheduling rules on the edges in order to minimize the sum of arrival times in a social optimum.

In the second part of the talk we assume that we are given some scheduling rules on the edges and discuss the inefficiency and existence of Nash equilibria depending on the chosen scheduling rules. We present some open questions that may be a start for future research.

3.12 Dynamic Atomic Congestion Games with Seasonal Flows

Marc Schroeder (Maastricht University, NL)

We propose a model of discrete time dynamic congestion games with atomic players and a single source-destination pair. The latencies of edges are composed by free-flow transit times and possible queuing time due to capacity constraints.

We give a precise description of the dynamics induced by the individual strategies of players and of the corresponding costs, either when the traffic is controlled by a planner, or when players act selfishly. Importantly, we model seasonalities by assuming that departure flows fluctuate periodically over time.

Our main contributions are two-fold. First, we introduce a measure that captures the queues induced by periodicity of inflows. For socially optimal flows, this measure is the increase in costs compared to uniform departures. The same holds for equilibrium flows, if the network is parallel. In general the analysis is more intricate. We even provide an example in which periodic departures induce lower equilibrium costs than the uniform departures. Second, we illustrate a new dynamic version of Braess's paradox: the presence of initial queues in a network may decrease the long-run costs in equilibrium. This paradox may arise even in networks for which no Braess's paradox was previously known.

3.13 Inefficiency caused by Risk Aversion in Selfish Routing

Nicolás E. Stier-Moses (Facebook – Menlo Park, US)

Joint work of Lianeas, Thanasis; Nikolova, Evdokia; Stier-Moses, Nicolás E.
Main reference E. Nikolova, N. E. Stier-Moses, "The Burden of Risk Aversion in Mean-Risk Selfish Routing," in Proc. of the 16th ACM Conf. on Economics and Computation (EC'15), pp. 489–506, ACM, 2015.
URL http://dx.doi.org/10.1145/2764468.2764485

Considering congestion games with uncertain delays, we compute the inefficiency introduced in network routing by risk-averse agents. At equilibrium, agents may select paths that do not minimize the expected latency so as to obtain lower variability. A social planner, who is likely to be more risk neutral than agents because it operates at a longer time-scale, quantifies social cost with the total expected delay along routes. From that perspective, agents may make suboptimal decisions that degrade long-term quality.

We define the price of risk aversion (PRA) as the worst-case ratio of the social cost at a risk-averse Wardrop equilibrium to that agents are risk-neutral. For networks with general delay functions and a single source-sink pair, we show that the PRA depends linearly on the agents' risk tolerance and on the degree of variability present in the network. In contrast to the price of anarchy, in general the PRA increases when the network gets larger but it does not depend on the shape of the delay functions. To get this result we rely on a combinatorial proof that employs alternating paths that are reminiscent of those used in max-flow algorithms. For series-parallel (SP) graphs, the PRA becomes independent of the network topology and its size. As a result of independent interest, we prove that for SP networks with deterministic delays, Wardrop equilibria maximize the shortest-path objective among all feasible flows.

3.14 Traffic signal optimization and user equilibria

Martin Strehler (TU Cottbus, DE)

Joint work of Strehler, Martin; Köhler, Ekkehard

We will present a cyclically time-expanded network model for simultaneously optimizing traffic assignment and traffic signals in an urban road network with a special focus on user equilibria. The model itself combines ideas from static and dynamic models. Being a completely linear approach, this allows the use of strict mathematical programming techniques. However, since a linear model may be too restrictive for real world applications, we investigate in particular travel times and latency functions. A comparison of travel times with simulation tools like MATSim and VISSIM demonstrates the applicability of our approach. These results give rise to numerous open questions concerning user equilibria.

3.15 New iterative algorithm for the Link Transmission Model: DTA designed in support of optimization procedures

Chris M. J.Tampère (KU Leuven, BE)

Main reference W. W. E. Himpe, R. Corthout, C. M. J. Tampère, "An efficient iterative link transmission model." Transporation Research Part B: under review (R1).
URL http://tinyurl.com/o65gbfc

Dynamic Traffic Assignment (DTA) models are often part of some bi-level optimization, e.g. calibration, optimal control, network design. If the optimization is to be solved iteratively, the DTA needs to be repetitively computed on the same network under marginally different inputs.

We present a DTA algorithm capable of calculating equilibria in medium to large networks consistent with LWR theory in feasible time. It is efficient especially when repeatedly calculating solutions on the same network, because both the Dynamic Network Loading (DNL) component and the Dynamic Shortest Path (DSP) calculation are done using a previously computed solution as a warm start. For the DNL this is achieved through a novel iterative algorithm. Rather than computing traffic states consistently at once with short (CFL-compliant) time steps, we iteratively compute a fixed point solution between forward propagation of multiple, destination-based commodity flows and back-propagation of congestion waves until consistency. The potential of the approach is illustrated on case studies of OD-estimation and anticipatory ramp metering, that take the DNL and DTA respectively as a constraint, the sensitivity of which is calculated numerically through finite differencing.

3.16 Dynamic network loading for congested networks in DTA: survey of theory, models and properties

Chris M. J. Tampère (KU Leuven, BE)

This presentation gives an overview of DNL as a supply model for DTA models. It shows the importance of the link and node models that are combined into the DNL. The DNL models in literature are categorized into a point-queue and a physical queue family, based on whether they only carry demand information downstream, or in addition also propagate supply constraints (congestion) upstream. Whereas this has been widely recognized for the link model, the role of an adequate node model is usually underestimated.

Node models combine the task of propagating consistently demand and supply constraints imposed by all connected link ends, with imposing internal constraints themselves (resulting from conflict points where traffic in different directions cannot simultaneously pass). It is shown that such node models exist, but they only have unique solutions under prohibitively strict conditions. The properties and behavior of physical queue models with state-of-the-art node models are illustrated using some pathological cases that have been described in the literature. These cases may have multiple stable solutions with transitions between them being triggered by simple events, they may exhibit oscillatory solutions even under stationary boundary flows, or may diverge to gridlock, a network state with zero flow and infinite travel time. The presentation concludes with an overview of properties of the travel time operator and DTA solution depending on the chosen DNL model.

3.17 Uniqueness of Nash Equilibria in Atomic Splittable Congestion Games

Veerle Timmermans (Maastricht University, NL)

Joint work of Timmermans, Veerle; Harks, Tobias

Congestion games where first introduced by Rosenthal in 1973 [1]. In these games a finite set of players competes over a finite set of resources. In a pure strategy every player chooses a subset of resources, and the cost for every resource depends on the number of other players choosing this resource. We consider the splittable variant, where every players has a positive demand, and in a pure strategy a player can split this demand over several strategies. Furthermore our costfunctions are player-specific.

We introduce the class of two-sided matching matroids, a class where every pair of bases contains a perfect matching in their symmetric difference. We prove that when the strategy space of every player consists of bases of a two-sided matching matroid, the Nash equilibrium is unique, no matter how these strategy spaces are interweaved. We show that this class contains laminar matroids, transversal matroids and graphic matroids of generalized series parallel graphs.

References

1 R. Rosenthal. *A class of games posessing pure-strategy Nash equilibria*. International Journal of Game Theory, 2(1):65–67 (1973).

3.18 Transportation Networks: Some Useful Tools for Analysing "Flows over time"?

Dave P. Watling (University of Leeds, GB)

Joint work of Watling, Dave P.; O'Hare, Steven; Connors, Richard; Sumalee, Agachai; Shepherd, Simon; Koh, Andrew; Rasmussen, Thomas; Nielsen, Otto; Prato, Carlo

The purpose of the presentation is to describe some issues arising in the field of static traffic network analysis (using non-atomic congestion games), which may be useful in analyzing problems in which network flows vary over time. Four issues are considered:

1. The Price of Anarchy (PoA) for such static games may vary considerably with a scaling of the origin-destination (s-t) demands, and the nature of this variation (e.g. in terms of the shape of relationship of PoA versus scale factor) differs across networks. Some recent research is briefly described which attempts to characterize "transition phases" in such a relationship (as the set of minimal cost paths or minimal marginal cost paths expands or contracts.
2. An alternative route choice model, representing travellers' mis-perceptions of travel costs/delays, which smoothly distributes flows over all permitted paths. This model is parameterized in a way that in one limit, the flows that arise approximate Wardrop equilibrium flows to an arbitrary accuracy.
3. Applications of this alternative route choice model for (a) analysing problems of network design (where the resulting mathematical program with equilibrium constraints is smooth in the design variables, given a smooth objective), and (b) modelling competitions between two cities in setting tolls, subject to travellers choosing route and whether to travel (an equilibrium problem with equilibrium constraints).
4. A second alternative route choice model in which relaxed conditions allow traffic to be distributed (in equilibrium) smoothly over only a sub-set of the permitted paths, with some paths unused.

3.19 Equilibrium Computation for Linear Complementarity Problems

Bernhard von Stengel (London School of Economics, GB)

The linear complementarity problem (LCP) generalizes linear programming (via the complementary slackness conditions of a pair of optimal primal and dual solutions) and finding Nash equilibria of bimatrix games. Lemke's classical complementary pivoting algorithm finds a solution to an LCP in many cases.

We give an exposition of Lemke's algorithm in geometric and algebraic terms, and explain the intrinsic local orientation of the computed path which is naturally defined in terms of signs of determinants. For traffic networks with affine delay functions, a Wardrop equilibrium is the solution to an LCP with the number of paths as its dimension. For details see [1].

References

1 Richard W. Cottle, Jong-Shi Pang, Richard E. Stone (1992), *The Linear Complementarity Problem.* Academic Press.

4 Open Problems

4.1 The Complexity of Computing a Maximum Robust Flow

Jannik Matuschke (TU Berlin, DE)

Consider a capacitated network with source s and sink t, and an integer k. Two players F and C play the following zero-sum game: First F specifies an flow on s-t-paths within the arc capacities, then C interdicts the flow by destroying a set S of k arcs. F's goal is to maximize the amount of surviving flow, i.e., the flow on paths that do not intersect S. For several years, computing the value of this very basic maximum flow interdiction game was believed to be NP-hard, even when k is fixed to 2. However, recently an error in the proof was discovered, and the complexity of the problem is open again.

4.2 Strategic or non-strategic blockages

Kai Nagel (TU Berlin, DE)

We have seen in several presentations, especially from the transport engineering side, that blockages play an important role, especially when they spill back across nodes. Yet, such models seem to be difficult to solve for the mathematicl approaches that are presented during the seminar. This presentation introduces some specific flow models, many of them explained in more detail by Chris Tampère in his review talk. The presentation then asks the question if it might be possible to come up with model classes and/or specific problem formations that might be more amenable to the tools that are currently discussed and/or if one could identify directions for progress.

The general direction of the presentation was prepared together with Ekkehard Köhler, Martin Strehler, Chris Tampère.

4.3 A polyhedral approach to thin flows

Neil Olver (VU University of Amsterdam, NL)

Koch and Skutella have given a structural characterization of equilibria in the deterministic queueing model, which aims to capture crucial dynamic aspects of traffic. They show that the equilibria flow can be decomposed into phases, and each phase can be described as a solution to a problem they dub “thin flows with resetting”. A number of fascinating questions remain open; in particular, it is is not known if thin flows with resetting can be computed in polynomial time.

Koch and Skutella give an efficient algorithm for a special case of the problem, “thin flows without resetting”, showing how it can be solved by solving multiple maximum flow problems. I will present an alternate viewpoint and show that thin flows without resetting can be seen as the optimal solution of a certain linear program. The (as of yet unrealized) hope is that this polyhedral perspective may be useful in attacking the general problem with resetting edges.

4.4 Selfish Routing with Uncertainty – The Role of Information

Alexander Skopalik (Universität Paderborn, DE)

We discuss open problems related to uncertainty and the role of information in selfish routing.

We consider a model of selfish traffic routing in which users are uncertain about the overall demand in the system. Instead users only have a belief (i.e., a probability distribution) about the current state of the system. Thus, users choose routes to minimize their expected cost.

We seek to understand the role of information about the state of the system on the performance of the system.
An easy example shows that equilibrium flow is improved if players have full information in comparison to having no information. We exhibit a – somewhat counter intuitive – example in which traffic flow is worse if users have information as opposed to the flow without information.

By combining the ideas of the two examples, we can show that there exist examples in which it is optimal to provide only partial information to the users.

This raises questions regarding the optimal signaling mechanism, its performance influence and computational complexity.

4.5 Open Question Session: "Flows over time" but not as we know it?

Dave P. Watling (University of Leeds, GB)

Joint work of Watling, Dave P.; Cantarella, Giulio E.
Main reference D. P. Watling, G. E. Cantarella, "Modelling sources of variation in transportation systems: Theoretical foundations of day-to-day dynamic models," Transportmetrica B: Transport Dynamics, 1(1):3–32, 2013.
URL http://dx.doi.org/10.1080/21680566.2013.785372

The focus of "flows over time" has been on the analysis of the kind of dynamics that would be observed within a day ("within-day dynamics"), exploring equilibrium properties arising from these time-dependent interactions. An alternative field of study in the transportation community is the study of "day-to-day dynamics", which describe the transient process of adaptations of travellers in their choices (which may or may not converge to some equilibrium). There are several classes of such models, and examples are given based on

1. a differential equation representing continuous-time, deterministic dynamics;
2. a difference equation representing discrete-time, deterministic dynamics;
3. a markov process representing discrete-time, stochastic dynamics.

The question is whether existing work in studying flows over time is extensible or could be adapted to understand the transient, non-equilibrium dynamics of such systems, in addition to the within-day dynamics conventionally.

Participants

- Umang Bhaskar
TIFR Mumbai, IN
- Roberto Cominetti
University of Chile – Santiago de Chile, CL
- José R. Correa
University of Chile – Santiago de Chile, CL
- Martin Gairing
University of Liverpool, GB
- Yannai A. Gonczarowski
The Hebrew University of Jerusalem, IL
- Tobias Harks
Maastricht University, NL
- Martin Hoefer
Universität des Saarlandes, DE
- Max Klimm
TU Berlin, DE
- Ekkehard Köhler
TU Cottbus, DE
- Felix König
TomTom – Berlin, DE
- Jannik Matuschke
TU Berlin, DE
- Kai Nagel
TU Berlin, DE
- Neil Olver
VU University of Amsterdam, NL
- Britta Peis
RWTH Aachen, DE
- Rahul Savani
University of Liverpool, GB
- Marco Scarsini
LUISS Guido Carli – Rome, IT
- Miriam Schlöter
TU Berlin, DE
- Daniel Schmand
RWTH Aachen, DE
- Marc Schröder
Maastricht University, NL
- Alexander Skopalik
Universität Paderborn, DE
- Martin Skutella
TU Berlin, DE
- Nicolás E. Stier-Moses
Facebook – Menlo Park, US
- Martin Strehler
TU Cottbus, DE
- Chris M. J. Tampère
KU Leuven, BE
- Veerle Timmermans
Maastricht University, NL
- Laura Vargas-Koch
RWTH Aachen, DE
- Bernhard von Stengel
London School of Economics, GB
- Dave P. Watling
University of Leeds, GB

Report from Dagstuhl Seminar 15421

Rack-scale Computing

Edited by

Babak Falsafi[1], Tim Harris[2], Dushyanth Narayanan[3], and David A. Patterson[4]

1 **EPFL – Lausanne, CH,** babak.falsafi@epfl.ch
2 **Oracle Labs – Cambridge, GB,** timothy.l.harris@oracle.com
3 **Microsoft Research UK – Cambridge, GB,** dnarayan@microsoft.com
4 **University of California – Berkeley, US**

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15421 "Rack-scale Computing". The seminar was successful and facilitated interaction between researchers working in a diverse set of fields, including computer architecture, parallel workloads, systems software, and programming language design. In addition to stimulating interaction during the seminar, the event led to a follow-on Workshop on Rack-Scale Computing to be organized during 2016.

Seminar October 11–16, 2015 – http://www.dagstuhl.de/15421
1998 ACM Subject Classification B.3 Memory Structures, C.1.4 Parallel Architectures, D.3 Programming Languages, D.4 Operating Systems
Keywords and phrases Rack-scale systems, Parallelism, Computer Architecture
Digital Object Identifier 10.4230/DagRep.5.10.35

1 Executive Summary

Babak Falsafi
Tim Harris
Dushyanth Narayanan
Kaveh Razavi

Rack-scale computing is an emerging research area concerned with how we design and program the machines used in data centers. Typically, these data centers are built from racks of equipment, with each rack containing dozens of discrete machines connected by Ethernet or by InfiniBand. Over the last few years researchers have started to weaken the boundaries between these individual machines, leading to new "rack-scale" systems. These architectures are being driven by the need to increase density and connectivity between servers, while lowering cost and power consumption.

Initial commercial systems provide high-density processor nodes connected through an in-machine interconnect to storage devices or to external network interfaces (e.g., HPE Moonshot, or SeaMicro Fabric Compute). Many ideas are now being explored in research projects – e.g., the use of custom system-on-chip processors in place of commodity chips, the use of emerging non-volatile-memory technologies or stacked Flash in place of disks, and the use of silicon photonics and wireless links for communication within or between rack-scale systems. In addition, researchers are exploring how systems software, language runtime systems, and programming models can evolve for these new architectures.

This seminar sought to bring together researchers working on different parts of these problems. We structured the seminar around a small number of invited introductory talks (Section 4) accompanied by break-out sessions (Section 5) and a series of four poster sessions. The poster sessions permitted everyone to have an opportunity to present their own work (if they wished to), and enabled many parallel discussions to continue at the same time around different posters.

2 Table of Contents

3 Overview of Talks

In the near future we expect to see rack-scale computers with 1000s of cores, terabytes of memory, high-bandwidth and low-latency internal fabrics. These new architectures raise several interesting research questions. What are the trade-offs between the different approaches being explored? How should the interconnect fabric be organized, and how should CPUs, DRAM, and storage be placed in it? Are different kinds of rack-scale systems required to handle different kinds of workloads efficiently? How should we integrate rack-scale computers into data center networks and warehouse-scale computers (WSCs)? To what extent should rack-scale machines be programmed as large shared-memory NUMA servers, or as traditional distributed systems, or a combination of the two? What are the correct communication primitives to let applications benefit from low-latency communication within the system? What are the likely failure modes and how do we achieve fault tolerance? How can researchers effectively prototype and test novel ideas?

Specifically, we sought to involve researchers and practitioners working on:

- Systems-on-Chip. SoCs are used both in industrial and research rack-scale systems, motivated by a drive for high density and high performance-per-Watt. For instance, Intel Atom C2000 processors (used in HPE Moonshot servers) provide up to eight 64-bit cores, a dual-channel memory controller, integrated PCIe, four GbE ports, along with SATA and USB. The UC Berkeley FireBox is a 50kW WSC building block containing a thousand compute sockets, each containing a SoC with around 100 cores connected to high-bandwidth on-package DRAM, and fast SoC network interfaces.
- Interconnects. A motivation for many rack-scale systems is to support workloads that benefit from more fine-grained communication than is possible between machines in traditional data centers. This includes developments at the physical level (for instance, silicon photonics). In addition, it includes development in networking, and in how the interconnect is exposed to software to enable high bandwidth communication or high message rates. For instance, the Scale-Out NUMA architecture exposes the interconnect via a remote memory controller which is mapped into a node's local cache-coherent address space. This removes the latency of crossing interfaces such as PCIe when communicating across a rack-scale system.
- Storage systems. There are many different technologies emerging for non-volatile random-access memory (NV-RAM), some promising high-capacity non-volatile storage coupled with read performance comparable with DRAM. At the same time, other researchers are exploring techniques for 3D-stacking FLASH and DRAM.
- Systems software and language runtime systems. How should operating systems handle rack-scale systems, and schedule work on them? Should a single operating system run over a complete system, as in today's large NUMA machines, or should separate operating systems run on each core, or each socket, or at some other granularity. How do we expose locality concepts, and what problems do we solve in hardware versus software. Experience with multi-core research operating systems, such as Barrelfish, fos, and Tessellation is relevant here. At a higher level, to what extent should rack-scale systems be programmed using the abstractions developed for distributed computing (such as replicated objects, actors, or message-passing), or via new implementations of the abstractions used for shared-memory (such as task-parallel programming models, or transactional memory). How are hardware failures, and tolerance of tail-latency accommodated in programming models?

The goal of this Dagstuhl research seminar was to bring together leading international researchers from both academia and industry working on different aspects of rack-scale systems. Effective solutions will require renewed collaboration across architecture, systems, and programming language communities. In addition, we sought to involve participants with practical experience of workloads, and of running industrial warehouse-scale systems.

4 Workloads and Technologies

In this report, we briefly summarize the seminar's introductory talks on some of the workloads and technologies relevant to rack-scale computing.

4.1 SpiNNaker

Steve Furber (University of Manchester, GB)

The SpiNNaker project aims to implement a massively-parallel system incorporating a million ARM processor cores for modeling large-scale systems of spiking neurons in biological real time. The 2D toroidal mesh must be folded into 10 racks with 3,600 high-speed interconnect cables, with no single cable longer than 1 meter. A purpose-built machine room with 100kW cooling capacity has been built to house the machine. This talk described the engineering challenges of constructing the machine to a modest budget.

Discussion notes

- Verifying that the hardware is doing what the spec says is a challenge (the hardware is non-deterministic).
- Power density and heat removal are also major challenges.
- Analogies between neuromorphic computing and a biological brain from a capability perspective are difficult. At best one can compare complexity/density of computational elements.

4.2 Data management at rack-scale

Gustavo Alonso (ETH Zürich, CH)

The killer application these days is big data, which translated to something practical means processing and managing large data collections of a variety of types (graphs, records, streams, tables, files, images, etc.). The platform of choice these days is a rack, either as a single unit or combined into a larger cluster. Racks are needed in the context of big data because of I/O and/or because of the need to support complex, concurrent workloads.

This talk focussed on these two aspects: I/O and workloads. Often forgotten in research papers, these are typically more important than CPU capacity and pose the real challenge in designing rack-scale data management systems.

4.3 Building Hardware for Microsoft's Datacenters

Leendert van Doorn (Microsoft Corporation – Redmond, US)

This talk briefly described some key workload criteria and design constraints for Microsoft's internal and external datacenters and how those shape our hardware designs. A key element of this is how to control the server cost. The talk specifically covered a set of new silicon technologies that are on the 3-6 year horizon that can help datacenters but also pose new questions. The talk ended with a case for disaggregated resource/rack scale designs and their challenges.

Discussion notes

- The design cycle for cloud infrastructure is roughly six months.
- There are two workloads of interest:
  1. First-party data analytics which leads to high utilization on bare metal.
  2. "Lift and shift" (virtualization) exhibiting low utilization but with QoS guarantees.
- Regarding supply chain, a key question is whether one size fit all in terms of compute/network/storage? Mostly yes, but premium hardware is useful for special workloads.
- Power is the limiting resource, not the compute density as per conventional wisdom. Racks are sometimes power-limited and not fully populated. Rack height is standardized (e.g., 48Us).
- The unit of failure is a rack. There is replication across clusters and geo-locations.
- How does the cost break down? 30% CPU, 70% the rest of the system (memory, SSD/HDD, networking, ...). DRAM price margins are so low that DRAM prices can not be publicized.
- Future: every SoC will include everything: cores, lots of memory, fast NICs, switches, and accelerators. There will be memory stacking and high-bandwidth serial links. How should one best take advantage of emerging memory technologies? How should we organize cache hierarchies and to what extent do we need NUMA?

4.4 Memory: Past, Present and Future Trends

Paolo Faraboschi (HP Labs – Palo Alto, US)

The talk surveyed the trend in memory technology and architecture, starting from the historical importance of DRAM in today's computing systems, all the way to future trends. It covers technology and business aspects that highlight the limitations and bottlenecks of the memory architecture we build in computing systems. The talk also covered the transition to non-volatile memory that is expected to happen in the near future, and the implication on the memory architecture organization.

Discussion notes

- Intel Developer Forum keynote 2000 – Quiz: Who was the guest who asked for more memory capacity to fit the whole web in DRAM? Answer: Larry Page.
- DRAM is the main medium for storing data. And over time the cost of memory has been going down constantly and reduced significantly. Zooming in the last few years (2000–today) we observe complete flattening of the cost of DRAM ($6-$10 per GB) and NAND ($0.3–$0.5 per GB).
- Memory is currently enslaved to the CPU memory channels; should we re-architect the way we place the memory? We need more system capacity, more chip density, bandwidth, byte addressability, more near-data intelligence, etc. Let's see some of the trends out there:
  - More bandwidth (beyond DDRx): serially attached Memory (Micron HMC). Two-tier applicability: (1) performance and (2) capacity. But currently the cost of HMC per GB is huge and therefore it is impractical to integrate it at scale.
  - Another alternative for getting more bandwidth is the 2.5D-3D integration. 2.5D packaging on Si interposer, 3D die stacking, or you can also go for system-in-package: combination of 3D die stacking and 2.5D interposer integration. Is this the solution? Well, yes, maybe. But stacking things like this (8x93%) results in a yield of 55%.
- There are three technology candidates (all NV): PC-RAM, STT-RAM, R-RAM and they all have different properties. The most important property is the density. If you don't get that one right, the cost is going to be huge and nobody is going to use it. And all these technologies are covering a wide-range of reading/writing latencies, and no energy cost for maintaining state (unlike DRAM). And of course these technologies have different maturity levels but we are not going to focus on that for now.
- There is another way to put it together: maybe I can put some memory in a fabric. Emancipated memory (outside of the control of the CPU), you have it in another fault domain and there are many other things you can do with it... But how much can one scale the memory fabric? What is the cost of the fabric? What about the latency?

4.5 Chips in Racks: What are the Trends?

Boris Grot (University of Edinburgh, GB)

The talk covered emerging trends in hardware for rack-scale computing. These trends include emergence of specialized many-core datacenter processors, pervasive accelerators, high-bandwidth memory technologies, and low-latency rack-scale fabrics. While individually, these technologies mitigate specific performance and energy-efficiency bottlenecks of today's systems, effectively leveraging the entire ensemble in a rack-scale deployment calls for potentially significant modifications to the system and application software stack.

Discussion notes

- Data is outgrowing compute power, 2% of world power is being spent in data centers.
- What is in a Xeon processors? Large LLC, few fat cores. LLC is not helpful for many workloads (100W).

Data center trends

1. Server processors.
   Q: Why have you married hw/sw w.r.t. coherence? Why not let the OS handle this?
   A: OS can if it wants; scale-out applications do not really need coherence
   Q: I wouldn't agree with the power argument (w.r.t. LLC). There is new technology that can replace LLC.
   A: These technology are for out of CPU.
2. Accelerators (GPU vs. FPGA) will help improve efficiency with the slowdown in Dennard Scaling.
3. Goodbye bandwidth wall – HMC is 10x faster than DDR3.
   Q: But HMC is small?
   A: That is not fundamental. we can have higher density in the future.
   Q: SerDes links may provide high bandwidth but the links need to be always on, and dissipate much power?
   A: Yes
4. Very fast intra-rack network – low energy.
   Q: But these photonics need transceivers (high energy)?
   A: No, laser on chip, so no need for transceivers. The audience was not unanimous about this.
5. Low-latency communication (integrated fabrics, Scale-out NUMA).
   Q: Point-to-point communication is not always the best for all workloads (e.g., graph processing). You want a lower hop-count due to the random nature of the workload.
   A: I agree, but these provide ample benefits for load-balancing.
   Q: I think having more hardware resources makes the problem on the application side harder. For example, it is really hard to write an application that can saturate 400 GB/s of bandwidth.
   A: Our point is that, changing the software to scale it with the number of cores is hard. Instead we argue that if you application runs fine with one pod of cores, by scaling the number of pods (i.e., partitioning the cores), the applications can scale easier without modification.
   Q: It is really hard to increase the memory bandwidth.
   A: We actually agree with that. That is why we argue for partitioning the bandwidth.

4.6 Rack-Scale Storage: Business as Usual or Something Different?

Ant Roswtron (Microsoft Research UK – Cambridge, GB)

This talk shed some light on what rack-scale computing means in general and what this means for storage specifically. The key is a converged rack-scale design – where all resources are carefully provisioned and managed to achieve a specific performance or price point. The talk used pubic cloud storage as the motivation, and MSR's experiences building out rack-scale storage for the cloud. The talk highlighted some ways forward and also pointed out some of the many pitfalls that the designers experienced.

Discussion notes

- Currently the unit of deployment is 1U-4U.
- Key principle: reduce $/GB given a certain performance – e.g., BlackBlaze backup: 4U, 45 HDD in JBOD.
- Latency (ms...hours) vs. Access pattern (hot – nearline – cold) vs. cost.
- Capacity vs. IOPS vs. latency vs. cost.
- Pelican: capacity for cold data (seconds latency) at a price point of tape. 8% of drives active at a time, power and cooling provisioned for those drives. Software enforces constraints: hard (power/cooling/vibration) and soft (bandwidth).
  Q: What is the difference between temperature between top to bottom of the rack?
  A: About 20 degrees, air comes at about 20C and goes out at 40C.
  Q: Effect of vibration due to fans?
  A: The way we do the layout we reduce the number of fun.
  Q: But there is also ambient noise?
  A: We have not seen any of that.
  Q: Is this a block interface for the rest of system?
  A: Think of it as blob service.
  Q: Rack is a unit of failure, this is not true here...
  A: We can fail over between servers and drives.
  Q: What percentage of storage would benefit from this?
  A: I think this is a significant fraction. Things like OneDrive, etc.
- Taming hardware was hard. ASIC designers like to add feature. We had to fix bugs in silicon that can take up to 6 months to fix. So we have to push complexity to software as much as possible. Hence we need tools for doing this.
  Q: Do you provide replication, availability inside the rack?
  A: Yes, we do some. But we have to think cross rack replication as well.
  Q: This is archival. So most of your bandwidth is going to serve writes
  A: We see usually multiple reads at a time.
  Q: Maybe you should do scheduling at multiple rack levels to reduce latency?
  A: Yes, that is a possibility.
  Q: What is the role of tiering in the rack-scale system? What is the proper way of moving things between tiers?
  A: We have to see whether we can have racks that can work in different tiers. If not, we have to design them based on a certain throughput.

5 Breakout Sessions

Much of the time during the seminar was devoted to breakout sessions. We briefly summarize the main topics of discussion which arose.

5.1 Research Challenges

Below, we describe rack-scale research challenges identified in different domains. This section is formed mostly in terms of open-ended questions or proposals.

Network Topology

- Should the interconnect be electrical, optical, or hybrid?
- It depends on the maturity of the technologies. How fast are these technologies going to mature?
- Which topologies should we use? What are the trade-offs that we want to make (latency vs. bandwidth)? Does the topology even matter?
- We need reasonable benchmarks for evaluating rack-scale networks.

Rack-Scale Memory

- What is the impact of NVM on the OS?
- Is it going to be more like a cluster or NUMA?
- We need to be able to control: the movement of data between volatile and non-volatile state and ordering of memory operations. Sync vs. async.
- Alternate/relaxed consistency – entry, release?
- Should systems use explicit messaging or implicit read/write?

Power

- Tradeoffs are between the cooling of rack and what goes inside the rack.
- Up to 20 KW per rack including cooling is reasonable today.

Rack Organization

- What is disaggregation?
  - Is memory close to processing?
  - Rack-poded memory ("emancipated") is attractive for large working sets. Is good latency/bandwidth possible?
  - Disaggregation ⇒ virtualization ⇒ Cost savings.
  - Trade-offs:
    - Cost of "infinitely fast" interconnects.
    - Cost of porting or reimplementing legacy apps.
- What is a server that I can buy?
  - Is it a node or a collection of nodes?
  - A node is a heterogeneous set of CPU/RAM/Storage/Network.
- Active memory components may help the disaggregation (e.g., RDMA).

Availability/Reliability

- What is the failure domain in a disaggregated world?
- MTBF for virtualized resources ≠ for physical resources.
- What about bit rot in DRAM/NVM?
- Byte-addressable NVM programming models remove the middleman ⇒ faulty programs.
- Replication is required for storage class reliability which increases the requirements on the interconnect bandwidth.

- Local versus global versus remote ⇒ Need remote copy to survive local failure.
- What is the role of OS? Authentication and authorization, setting up mappings, sharing, etc.

Security

- We need enclave-like approaches to encrypt everything but the CPU caches.
- The security of new operating systems, file systems, etc. needs to be verified.
- The identity of (disaggregated) resources needs to be verified to avoid spoofing.
- Fine-grained protections versus course-grained address spaces. Lessons learned with CHERI using variable-length segments protected through capabilities are applicable here.
- What are the OS integration issues with capabilities?
- End-to-end security for rack-scale.
  - Compute, network, storage.
  - Key management.
  - Multi-tenancy.

5.2 Research Tools for Rack-Scale Computing

Benchmarks

We need new benchmarks and evaluation methodologies for rack-scale computers. The evaluation requires workloads over small and large datasets.

Debugging

For debugging, we will likely still rely on logging. On top of that, support for tracing and trace inspection will be instrumental. In addition, we require support from hardware for synchronized clocks and to able to inspect remote memory, tap into the network, and replay. We also need to build tools to identify coherence errors across nodes, deadlocks, live locks, lost packets, and messages.

Performance

We require additional information on top of what commodity hardware and networks provide. For example, Intel PT-like tracing can allow us to understand performance behavior of rack-scale applications. We need to be able to identify the number of hops a packet traverses, traffic information between A and B, and in general be able to query for locality information.

Simulation

Rack-scale computers can be simulated at the different levels. There are possibilities for architecture simulation as well as network simulation. FPGAs are a good test-bed for running these simulations that involve a large number of events. Further, rack-scale computers can be emulated. Emulation of slowness is possible by adding delays, and emulation of fastness is possible by making everything else in the system slower. We require test-bed prototypes for rack-scale research. We either need to build proofs or use small machines and project the results to larger machines.

5.3 Rack-Scale Networking

There are certain advantages to rack-scale networks. They allow for significant cost savings by allowing resource disaggregation. Further, they make it easy to ship new technology inside the rack and hiding it. How is the rack seen from the rest of the world? Is it going to be an open-box or a black-box, a single computer or a collection of computers? The rack is already a shipping unit. Note that the answer to these questions has implication for naming and failure domains.

Do we need to have two networks (one for management tasks) or is one enough? Also, another important aspect is the choice low-radix versus high-radix switches. Low radix switches (potentially embedded in SoCs) are better in terms of economy of scale but have higher latencies.

5.4 Architectural Support

Synchronization

We require atomically ordered multicast for implementing locks, transactions, etc. It can be implemented as a separate bus in a rack-scale computer. Spinning should be banned, but we need a high-performance wakeup.

Communication

The rack network needs to support multicast, provides high packet injection rate with low latency and high throughput. Connection-less RDMA or migratable RDMA connections to scale RDMA communication is desired. We need to develop efficient routing and congestion control inside the rack. The network needs to support virtualization for handling multiple users and migration. It further needs to provide programmable wake-ups on various events.

Fault tolerance

The rack needs to forward error correction on communication channels. It further needs to provide support for observing error rate counters and allow for isolation of components. We need support for replication and migration of memory in face of failures. Failure detection and/or prediction can be implementation as a stand-alone service.

Caching

Software needs control over local caches. Rack nodes need to provide support for cache specific load instructions to potentially more than one L1 cache (similar to scratch-pads). Variable-sized cache-lines and programmable prefetch engines are also desired in a rack-scale computer.

Accelerators

Accelerators allow for heterogeneous processing that allows for power efficiency. These accelerators need to support multiple processes. Accelerators that reduce data movement (i.e., near memory processing) and allow for programmable I/O are desired. Integrated FPGAs on the sockets can bring flexibility for processors in a rack.

QoS

The rack should provide strong QoS throughout the rack in order to meet SLAs. Introspection support will allow finding hot-spots using performance counters.

5.5 Operating Systems

The OS should provide the following set of services for a rack-scale computer:
- Failure detection.
- Location-transparent communication.
- Resource allocation/QoS.
- State machine replication as a service.

The hardware can provide these services to make it easier for the OS mechanisms:
- Warning before failure.
- Clearly defined "fault containers": memory, threads, etc. but also at the language level.
- Are these just transactions?
- Turn suspected failure into fail stop.

Resource allocation: Disaggregation gives great opportunities for scheduling/migrating/consolidation/dealing with failures/shared resources. But it adds complexity. Can we have a single level of scheduling? Is there anything we can do to make it simpler? Hierarchical versus "simple" scheduling. Can we do traditional OS tasks centralized or not? This generalizes to all the resources and sharing of all resources.

Can we do sharing via exclusive ownership of state? Partition and migrate the state and the ownership. The challenge is how to quickly and correctly change ownership. Shared/global abstractions with different implementations depending on the hardware? If IaaS, then OS layer can be very simple. If PaaS then containers? Either way, performance isolation and QoS are key. Memory bandwidth will be a new challenge to provide QoS when it is dis-aggregated. Scale will also make it harder.

Communities need to work together: we are talking to each other. Should we replicate functionality? Observation: we do already replicate functionality, can we get rid of this (VMM, OS, JVM). Why do you need to virtualize at all? Should we just do containers? A lot of people liked the idea of exporting info between DB and OS and doing affinity scheduling.

Resource monitoring on a large scale. How do you debug it? It's a distributed system. Not just globally but within each application.

5.6 High Performance Computing

HPC versus general-purpose computing

HPC systems resemble rack-scale computers to some extent. 99% of rack-scale computers are built to run multi-purpose applications and their scale is often a few large machines. HPC systems are rarely a success except for science for which they are built for, but the technology of the ones that are successful are transferred into general-purpose computing. Examples include GPUs, RDMA, and fat-tree topologies.

Programming model

MPI is often used as a library to program communicating applications on a HPC system. MPI is not a programming model, but provides a communication abstraction analogous to sockets. It is similar to assembly in which many programming models have been built on top of it. These models, however, are slow compared to native MPI, hence everyone instead uses MPI natively.

MPI is slowly moving away from message passing. One-sided put/get operations combined with fences allow for efficient transfer of data. You can think of it as a new ISA for programming HPC systems. There exists an axiomatic formal memory model that provides a very loose consistency model.

Reliability/Faults

In the HPC world, when faults happen the application is either started from the beginning or resumed from a checkpoint if the system supports checkpoints. Instead, in the data center world partial failures in stateful applications happen. ACID, Paxos, and alike are developed to deal with this. It would be interesting to investigate what happens with libfabric (a commonly used HPC library for communication) in face of partial failures.

Topologies

HPC networks focus on cheaper topologies. The main metric is the total global throughput on random uniform traffic. Low diameter topologies such as Dragonfly where groups of nodes are fully connected with full connection between groups perform well with this metric. These topologies however are not very modular: we cannot connect extra nodes with local changes, so the data center model of "wheel in an extra container and plug in electricity/water/power" does not directly apply.

Participants

- Gustavo Alonso
ETH Zürich, CH
- Yungang Bao
Chinese Academy of Sciences – Beijing, CN
- Angelos Bilas
FORTH – Heraklion, GR
- Peter Corbett
NetApp – Sunnyvale, US
- Paolo Costa
Microsoft Research UK – Cambridge, GB
- Christina Delimitrou
Stanford University, US
- Felix Eberhardt
Hasso-Plattner-Institut – Potsdam, DE
- Lars Eggert
NetApp Deutschland GmbH – Kirchheim, DE
- Babak Falsafi
EPFL – Lausanne, CH
- Paolo Faraboschi
HP Labs – Palo Alto, US
- Christof Fetzer
TU Dresden, DE
- Steve Furber
University of Manchester, GB
- Jana Giceva
ETH Zürich, CH
- Matthew P. Grosvenor
University of Cambridge, GB
- Boris Grot
University of Edinburgh, GB
- Hermann Härtig
TU Dresden, DE
- Tim Harris
Oracle Labs – Cambridge, GB
- Maurice Herlihy
Brown Univ. – Providence, US
- Matthias Hille
TU Dresden, DE
- Torsten Hoefler
ETH Zürich, CH
- Konstantinos Katrinis
IBM Research – Dublin, IE
- Kimberly Keeton
HP Labs – Palo Alto, US
- John Kim
KAIST – Daejeon, KR
- Christoph M. Kirsch
Universität Salzburg, AT
- Sergey Legtchenko
Microsoft Research UK – Cambridge, GB
- Martin Maas
University of California – Berkeley, US
- Sue Moon
KAIST – Daejeon, KR
- Andrew W. Moore
University of Cambridge, GB
- Dushyanth Narayanan
Microsoft Research UK – Cambridge, GB
- Jörg Nolte
BTU Cottbus, DE
- Mark H. Oskin
University of Washington – Seattle, US
- Simon Peter
University of Washington – Seattle, US
- Andreas Polze
Hasso-Plattner-Institut – Potsdam, DE
- Danica Porobic
EPFL – Lausanne, CH
- Zoran Radovic
Oracle – Stockholm, SE
- Kaveh Razavi
VU University Amsterdam, NL
- Randolf Rotta
BTU Cottbus, DE
- Ant Rowstron
Microsoft Research UK – Cambridge, GB
- Stefan Schmid
TU Berlin, DE
- Bernhard Schräder
Fujitsu Technology Solutions GmbH – Paderborn, DE
- Malte Schwarzkopf
MIT – Cambridge, US
- Liuba Shrira
Brandeis Univ. – Waltham, US
- Jens Teubner
TU Dortmund, DE
- Gael Thomas
Télécom & Management SudParis – Evry, FR
- Jana Traue
BTU Cottbus, DE
- Leendert van Doorn
Microsoft Corporation – Redmond, US
- Haris Volos
HP Labs – Palo Alto, US
- Bernard Wong
University of Waterloo, CA
- Noa Zilberman
University of Cambridge, GB
- Ferad Zyulkyarov
Barcelona Supercomputing Center, ES

Report of Dagstuhl Seminar 15431

Genomic Privacy

Edited by

Jean Pierre Hubaux[1], Stefan Katzenbeisser[2], Bradley Malin[3], and Gene Tsudik[4]

1 **EPFL – Lausanne, CH,** jean-pierre.hubaux@epfl.ch
2 **TU Darmstadt, DE,** skatzenbeisser@acm.org
3 **Vanderbilt University – Nashville, US,** b.malin@vanderbilt.edu
4 **University of California – Irvine, US,** gts@ics.uci.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15431 "Genomic Privacy". The current rise of personalized medicine is based on increasing affordability and availability of individual genome sequencing. Impressive recent advances in genome sequencing have ushered a variety of revolutionary applications in modern healthcare and epidemiology. In particular, better understanding of the human genome as well as its relationship to diseases and response to treatments promise improvements in preventive and personalized healthcare. However, because of the human genome's highly sensitive nature, this progress raises important privacy and ethical concerns, which simply cannot be ignored. A digitized genome represents one of the most sensitive types of human (personal) identification data. Even worse, a genome contains information about its owner's close relatives. The Dagstuhl seminar 15431 brought together computer scientists, bioinformaticians, geneticists and ethical experts to discuss the key security and privacy challenges imposed by the storage of large volumes of genetic data.

Seminar October 18–23, 2015 – http://www.dagstuhl.de/15431
Keywords and phrases cryptography, differential privacy, genetics, genomics, health data, information security, privacy by design, privacy protection, secure computation
Digital Object Identifier 10.4230/DagRep.5.10.50

1 Executive Summary

Jean Pierre Hubaux
Stefan Katzenbeisser
Bradley Malin
Gene Tsudik

This report documents the program and the outcomes of Dagstuhl Seminar 15431 "Genomic Privacy". The current rise of personalized medicine is based on increasing affordability and availability of individual genome sequencing. Impressive recent advances in genome sequencing have ushered a variety of revolutionary applications in modern healthcare and epidemiology. In particular, better understanding of the human genome as well as its relationship to diseases and response to treatments promise improvements in preventive and personalized healthcare.

At the same time, human genetics has become a "big data" science. For roughly a decade, specific tests for Single Nucleotide Polymorphisms (SNPs), e.g., markers corresponding to specific diseases, have been well established. Furthermore, research in pharmaco-genomics, which currently relies on SNPs, has helped improve drug treatment for cancer and cardiac patients. The methodology of genotyping, which takes into account hundreds to thousands of variations in positions in the genome, has tremendously increased the amount of data acquired during diagnosis. Personalized genotyping has become commercially available from several sources (such as 23andMe). Full genome sequencing and genome-wide association studies are moving towards full deployment in clinical practice. In 2000, the cost of sequencing one human genome was US$2.5 billion. Today, the price of US$200 for genome sequencing is approaching reality. Considering the benefits for (public) health and potential cost savings, widespread acquisition, storage, and usage of personal genomes is guaranteed to happen soon.

However, because of the human genome's highly sensitive nature, this progress raises important privacy and ethical concerns, which simply cannot be ignored. A digitized genome represents one of the most sensitive types of human (personal) identification data. Even worse, a genome contains information about its owner's close relatives. Furthermore, correlations with individual data sets from so-called "omics-technologies" pose even bigger threats on privacy. Leakage of personal genomic information can lead a wide variety of attacks, many of which are not yet fully understood. Whether accidentally or intentionally revealed, a digitized genome cannot be revoked or modified. Consequently, secrecy of personal genomic data is of paramount importance. Furthermore, genomic data, unlike other types of highly sensitive information (even national secrets), does not lose its sensitivity over time. Even worse, the mechanisms available to interpret genomic data improve over time, which means that it is unclear at the moment how much sensitive information a genome encodes and which consequences a genomic data breach has. Furthermore, it is likely that genomic data will not only be used personally to support medical treatments; great promise lies in its use in large-scale genetic studies for personalized medicine as well as common ancestry and genetic compatibility tests. Therefore, simply encrypting genomic data at rest is not a viable option and new ways of protection need to be devised.

The second Dagstuhl Seminar on Genomic Pricacy concentrated on the following topics:

- Technical solutions for genomic privacy: the participants discussed technical solutions to enable genomic data privacy, even in the presence of untrusted computing environments, and investigated technical protection techniques that can be used for this purpose.
- Integration of genomic and physiological data: For medical purposes, genomic data often needs to be correlated with clinical and physiological data. For example, clinical studies may require finding correlations between physiological data reported during hospital stays and genomic information. So far, most technical solutions for the protection of genomic data focused on securely storing DNA data itself, but did not discuss the complex problem of combining it with physiological data.
- Protection of sensitive data within large-scale genome-wide association studies: Although large-scale genomic studies offer many advantages for medical research, they pose many privacy problems. Most prior technical solutions focus on protection of a single human genome and do not scale multitudes of genomes. It remains a challenge to devise scalable techniques.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Privacy in the Genomic Era

Erman Ayday (Bilkent University – Ankara, TR)

Genome sequencing technology has advanced at a rapid pace and it is now possible to generate highly-detailed genotypes inexpensively. The collection and analysis of such data has the potential to support various applications, including personalized medical services. While the benefits of the genomics revolution are trumpeted by the biomedical community, the increased availability of such data has major implications for personal privacy; notably because the genome has certain essential features, which include (but are not limited to) (i) an association with certain diseases, (ii) identification capability (e.g., forensics), and (iii) revelation of family relationships. Moreover, direct-to-consumer DNA testing increases the likelihood that genome data will be made available in less regulated environments, such as the Internet and for-profit companies. The problem of genome data privacy thus resides at the crossroads of computer science, medicine, and public policy. While the computer scientists have addressed data privacy for various data types, there has been less attention dedicated to genomic data. Thus, the goal of this paper is to provide a systematization of knowledge for the computer science community. In doing so, we address some of the (sometimes erroneous) beliefs of this field and we report on a survey we conducted about genome data privacy with biomedical specialists. Then, after characterizing the genome privacy problem, we review the state-of-the-art regarding privacy attacks on genomic data and strategies for mitigating such attacks, as well as contextualizing these attacks from the perspective of medicine and public policy. This paper concludes with an enumeration of the challenges for genome data privacy and presents a framework to systematize the analysis of threats and the design of countermeasures as the field moves forward.

3.2 Efficient Server-Aided Secure Two-Party Function Evaluation with Applications to Genomic Computation

Marina Blanton (University of Notre Dame, US)

Computation based on genomic data is becoming increasingly popular today, be it for medical or other purposes such as ancestry or paternity testing. Non-medical uses of genomic data in a computation often take place in a server-mediated setting where the server offers the ability for joint genomic testing between the users. Undeniably, genomic data is highly sensitive, and there is an urgent need to protect it, especially when it is used in computation for what we call as recreational non-health-related purposes. Towards this goal, in this work we put forward a framework for server-aided secure two-party computation with the security model motivated by genomic applications. One particular security setting that we treat in this work provides stronger security guarantees with respect to malicious users than the traditional malicious model. In particular, we incorporate certified inputs into secure computation based on garbled circuit evaluation to guarantee that a malicious user is unable to modify her inputs in order to learn unauthorized information about the other user's data.

3.3 Privacy-preserving bioinformatics with general-purpose SMC

Dan Bogdanov (Cybernetica AS – Tartu, EE)

Research has given us a range of privacy technologies. From simple, yet not provably secure, technologies like pseudonymization to full encrypted processing with homomorphic cryptography or garbled circuits. Other technologies hide privacy by adding noise – anonymization and differential privacy are the main examples.

Each such technology has different trust, deployment and performance guarantees that are not immediately clear without a risk-benefit analysis. For example, statistical anonymization can often be defeated by linking with auxiliary information. Some homomorphic encryption schemes do not allow data to be collected from multiple owners and linear secret sharing leads to the best-performing programmable secure computing schemes.

Our work has focused on implementing a range of genomic analyses using secure multiparty computation based on secret sharing. We build our work on the Sharemind framework that supports integer, fixed point, floating point and boolean arithmetic and is easily programmable using the SecreC programming language.

After successfully demonstrating secure genome-wide association studies on Affymetrix microarray data with 500K SNP locations on 1000 patients we started developing a full statistical analysis system.

The Rmind privacy-preserving statistical tool is designed to mimic the popular R statistical tool. Rmind supports filtering, a range of statistical tests and also allows for corrections when many parallel tests are performed simultaneously. Rmind performs all operations using special data-independent algorithms that do not leak private inputs through the running time. Combining this with a secure computing environment gives us unparalleled privacy guarantees. Most recently, in 2015, we were able to add Principal Component Analysis to the list of operations supported on Sharemind.

We believe that by making Sharemind support different deployment models through the use of various secure computing protocols, we can further expand its usability in privacy-preserving personalized medicine.

3.4 Genomic privacy in research and medicine: a view from the trenches

Jacques Fellay (EPFL – Lausanne, CH)

There is a need to reconcile technical and theoretical development in genomic privacy with the reality of the research and medical worlds. Using concrete examples, I will describe current applications of genomics, at the crossroad between academic research and personalized medicine.

3.5 GenoGuard: Protecting Genomic Data Against Brute-Force Attacks

Zhicong Huang (EPFL – Lausanne, CH)

Secure storage of genomic data is of great and increasing importance. The prevalent use of passwords to generate encryption keys poses an especially serious problem when applied to genetic data. Weak passwords can jeopardize genetic data in the short term, but given the multidecade lifespan of genetic data, even the use of strong passwords with conventional encryption can lead to compromise. We present a tool, called GenoGuard, for providing strong protection for genomic data both today and in the long term. We prove that decryption under any key will yield a plausible genome sequence, and that GenoGuard offers an information-theoretic security guarantee against message recovery attacks. We also explore attacks that use side information. Finally, we present an efficient and parallelized software implementation of GenoGuard.

3.6 Efficient privacy-preserving deterrence of inference attacks on genomic data

Florian Kerschbaum (SAP SE – Karlsruhe, DE)

Many methods are known to privately query, analyze and compare genomic data. However, as has been shown by Goodrich in the Mastermind attack, repeated queries leak sufficient information in the result in order to quickly infer the secret genomic data. In this talk I will present a method that can deter such attacks in an efficient and secure manner using fuzzy commitments and zero-knowledge proofs.

3.7 Reality check – Implementing personalized therapies based on genomic data in a clinical setting

Oliver Kohlbacher (Universität Tübingen, DE)

Personalized immunotherapies based on epitope-based vaccines are an exciting strategy for personalized cancer treatment. The integration of different types of high-throughput data (exome, transcriptome, proteome, HLA ligandome) poses a number of interesting research problems. Implementing this into a clinical setting, however, results in several regulatory, organisational, and legal issues that we will discuss in the context of the iVac project implemented for personalized cancer immunotherapy at the university hospital in Tübingen.

3.8 Challenges faced by Hospitals using NGS for Diagnostics

Adam Molyneaux (Sophia Genetics SA – Lausanne, CH)

The presentation will talk about the practical problems faced by hospitals in using NGS as a diagnostic tool, explaining how Sophia set out to help them, where we are now and where we think we need to go in the future.

3.9 Controlled Functional Encryption

Muhammad Naveed (University of Illinois – Urbana-Champaign, US)

U.S. Department of Health & Human Services reports that health records of more than 39 million individuals have been breached from hospitals and other healthcare institutions. Therefore, a patient may worry about the privacy of her sensitive health data, and may want healthcare providers to learn only limited information, e.g., result of a particular test. In general, a patient may want healthcare providers to only learn information allowed by the patient specified policy. Such privacy concerns are not limited to healthcare domain. Existing cryptographic techniques do not provide realistic solution to this problem. For example, secure computation would require each patient, or an agent of the patient, to run a computationally intensive program on her computer for each computation. Functional encryption can solve the problem, but it is extremely inefficient and is based on untested cryptographic hardness assumptions.

In this work, we propose a new cryptographic model called "Controlled Functional Encryption (C-FE)" that allows us to construct realistic and efficient constructions. As in functional encryption, C-FE allows a user (client) to learn only certain functions of encrypted data, using keys obtained from an authority. However, we allow (and require) the client to send a fresh key request to the authority every time it wants to evaluate a function on a ciphertext. We propose two C-FE constructions: one for inner-product functionality and other for any polynomial-time computable functionality. The former is based on careful combination of CCA2 secure public-key encryption with secret sharing, while later is based on careful combination CCA2 secure public-key encryption with Yao's garbled circuit. Our main contributions in this work include developing and formally defining the notion of C-FE; designing efficient and practical constructions of C- FE schemes achieving these definitions for specific and general classes of functions; and evaluating the performance of our constructions on various application scenarios.

Our constructions are based on efficient cryptographic primitives and perform very well in practical applications. On a laptop, with Intel Core i7 processor and 8GB RAM, our construction takes 1.28s and consumes 132KB bandwidth for a 1,000 SNP disease marker in personalized medicine application. In genomic patient similarity application, comparing two 4-million SNP profiles costs $0.0143, takes 4 minutes and consumes 53.77MB bandwidth.

3.10 Engineering data privacy – The ARX data anonymization tool

Fabian Prasser (TU München – Klinikum Rechts der Isar, DE)

One of the main focuses of the seminar are privacy problems that arise from the combination and correlation of genomic and clinical data. While privacy for genomic data does pose significant challenges, problems with privacy of clinical data must not be underestimated. While a plethora of methods have been proposed for dealing with many aspects of de-identifying such data, only few (prototypical) implementations are available. Actually, the complexity of implementing privacy technologies is an often overlooked challenge.

In this talk we will present the open source anonymization tool ARX, which has been carefully engineered to support multiple privacy technologies for relational datasets. Our tool bridges the gap between different scientific disciplines by integrating methods developed and used by the statistics community with data anonymization techniques developed by computer scientists. ARX has been designed from the ground up to ensure scalability and it is able to process very large datasets on commodity hardware. The software implements a large set of privacy models: (1) syntactic privacy models, such as k-anonymity, l-diversity, t-closeness and d-presence, (2) statistical models for re-identification risks, and (3) differential privacy. Moreover, it supports multiple risk models and more than ten different methods for evaluating data utility, including loss, precision, non-uniform entropy and KL divergence. Data can be transformed automatically, semi-automatically and manually using a complex method that integrates global recoding, local recoding, categorization, generalization, suppression, microaggregation and top/bottom-coding. All methods are accessible via a comprehensive cross-platform graphical user interface.

Our talk will contribute to the overall seminar topic by comprising an overview of the possibilities and limitations of modern anonymization tools. We will also discuss challenges and possible further developments.

3.11 On a Novel Privacy-Preserving Framework for Both Personalized Medicine and Genetic Association Studies

Jean-Louis Raisaro (EPFL – Lausanne, CH)

So far, several efforts have been undertaken for protecting genomic data and still enabling its functionality. We can put them in two distinct categories: (i) approaches for private clinical genomics, and (ii) approaches for privacy-preserving genetic research. Yet, a main limitation of these approaches is that they restrict the private use of the data only to a single specific purpose, thus significantly slowing down the deployment of privacy-enhancing technologies in a real operational setting. In this work, we address this limitation by proposing a new privacy-preserving framework that is flexible enough to enable for both personalized medicine and genetic association studies on encrypted patients' data. Based on our previous research on private disease risk tests, we extend the previously proposed system model proposed in order to support also privacy-preserving replication and fine-mapping genetic association studies under the assumption of an honest-but-curious adversary. In particular, patients' data are stored encrypted on a centralized storage and processing unit (SPU), and the different healthcare stakeholders, or medical units (MU), can only obtain the study end-result without ever seeing the actual data.

3.12 Applying Homomorphic Encryption for Practical Genomic Privacy

Kurt Rohloff (NJIT – Newark, US)

This talk outlines a lattice encryption scheme that provides Proxy Re-Encryption (PRE) capabilities with Homomorphic Encryption (HE). We identify several high-level use cases for a mixed PRE and HE capability. We discuss early and implementation and experimental results.

3.13 Realizing differentially private genome-wide association studies

Sean Simmons (MIT – Cambridge, US)

The growing stockpiles of genomic data found in biomedical repositories and patient records promise to be an invaluable resource for improving our understanding of human diseases. In particular, there is interest in using this genomic data to perform genome wide association studies (GWAS). Recent work, however, has shown that sharing this data–even when aggregated to produce p-values, regression coefficients, or other study statistics–may compromise patient privacy.

One proposed solution is to use a privacy preserving technique known as differential privacy. This approach, which works by slightly perturbing the data, protects patient privacy while still allowing researchers access to their genomic data. Unfortunately, existing differentially private GWAS techniques have limitations in terms of accuracy, computational efficiency, and their ability to deal with heterogeneous populations. In this presentation I will give an overview of recent work we have done to help overcome these bottlenecks, work which moves privacy preserving GWAS closer to real world applicability.

3.14 Robust Traceability from Trace Amounts

Adam Davison Smith (Pennsylvania State University – University Park, US)

The privacy risks inherent in the release of a large number of summary statistics were illustrated by Homer et al (PLoS Genetics, 2008), who considered the case of SNP allele frequencies obtained in a genome-wide association study: Given the minor allele frequencies from a case group of individuals diagnosed with a particular disease, together with the genomic data of a single target individual and statistics from a sizable reference dataset independently drawn from the same population, an attacker can determine with high confidence whether or not the target is in the case group.

In this work we describe and analyze a simple attack that succeeds even if the summary statistics are significantly distorted, whether due to measurement error or noise intentionally introduced to protect privacy. Our attack only requires that the vector of distorted summary statistics is close to the vector of true marginals in ℓ_1 norm. Moreover, the reference pool required by previous attacks can be replaced by a single sample drawn from the underlying population. The new attack, which is not specific to genomics significantly generalizes recent lower bounds on the noise needed to ensure differential privacy (Bun, Ullman, and Vadhan, STOC 2014; Steinke and Ullman, 2015), obviating the need for the attacker to control the exact distribution of the data. In particular, the attack shows that natural relaxations of differential privacy (such as "Pufferfish", "coupled-worlds privacy" and related notions) are subject to the same lower bounds as full-strength differential privacy when many one-way marginals are released.

3.15 e-Biobanking: architectural and algorithmic solutions for the seamless, safe and secure storage and sharing of large biomedical data

Paulo Jorge Veríssimo (University of Luxembourg, LU)

The biomedical data lifecycle is changing dramatically, both due to factors like the generalization of physical sample collection, or the advent of NGs machines, and also due to the pressure for data sharing in name of the progress of biomedical research. Both factors have been inducing ad-hoc technical solutions, bolted on the classical lifecycle, such as use of clouds and promotion of web access, which are bound to augment the threat surface in non-negligible ways. We have been researching on avenues which preserve the desired functional evolution, but satisfy the need for built-in, by-design privacy, integrity and availability of such critical data. This talk reports advances toward new distributed systems architectures and privacy-preserving algorithms which, if successful, may foster what we call e-biobanking ecosystems, coalitions of stakeholders including hospitals, researchers, biobanks, or NGS providers.

Drawing on recent results based on the cloud-of-clouds paradigm, we first show how innovative distributed architectures may foster the advent of secure and dependable constellations of private and public clouds belonging to diverse stakeholders, with separation of risk and concerns, for example, making researchers able to perform operations on mix-criticality data residing in public and private clouds. Secondly, we show how to prevent concrete re-identification attacks on genomic data, leveraging on the above-mentioned architectural framework. We propose a method that systematically detects privacy-sensitive DNA segments coming directly from an input stream. Our method neutralizes threats related to recently published attacks on genome privacy based on short tandem repeats, disease-related genes, and genomic variations. The method can evolve automatically as new privacy-sensitive sequences are identified. Furthermore, the detection machine easily fits the e-biobanking model, by streamlining the cloud storage with the NGS production cycle by using Bloom filters and scaling out to faster sequencing machines.

3.16 The NIH Genome Privacy Challenges: Bringing Security Technologies to Biomedical Users

Xiao Feng Wang (Indiana University – Bloomington, US)

The growth of genome data and computational requirements overwhelm the capacity of servers. Many institutions and NIH are considering the cloud computing service as a cost-effective alternative to scale up research. Privacy and security are the major concerns when deploying cloud-based data analysis tools. In the past few years, progress has been made on secure data-dissemination and computation technologies but it is still not clear the gap between what they can provide and what are expected in the biomedical community. In the past two year, the genome privacy team at Indiana University works together with the iDASH NCBC center organized two NIH-sponsored genome privacy competitions. In this talk, I will provide information about these challenges and what we have learnt.

4 Working groups

4.1 Data sharing across domains

Emiliano De Cristofaro (University College London, GB)

This working group focused on understanding the requirements and the objectives of genomic data sharing in the research environment vis-a-vis the related privacy and security challenges it poses. In particular, participants analyzed existing initiatives as researchers have already developed protocols for exchanging DNA information across the web. Initiatives like the Global Alliance and Matchmaker Exchange1 aim to make it easier for geneticists and bioinformaticians to search, share, and retrieve genomic and epigenomic data that can help them with their research as well as treatment experimentation across institutions. However, while these projects put forward self-regulated codes of conduct and frameworks guided by human rights principles, non-discrimination, and procedural fairness, ultimately, their privacy practices boil down to reliance on volunteers' informed consent as well as ethical guidelines punishing misuse, intentional de-anonymization, or wide disclosure of personally identifiable information. The working group concluded that closer collaborations and exchanges need to take place so that privacy can be embedded from the outset in these protocols. In particular, participants agreed that data minimization approaches should be followed, without requiring the presence of fully-trust parties or shifting the liability of data leaks on the researchers. To this end, solutions from cryptography and differential privacy can offer viable promising opportunities but a number of research problems remain open with respect to efficiency, scalability, resilience to errors, and the lack of off-the-shelf readily available to non-cryptography experts.

4.2 Architecture and middleware

Aniket Kate (Purdue University – West Lafayette, US)

This working group focused on genomic data integrity issues and middleware architectures to make genome processing system transparent to involved clientele. With its heterogeneous nature, mixed criticality, and longevity and irrevocability issues, genomic data presents unique challenges in terms of integrity, privacy and reliability. It is observed that we cannot ignore arbitrary faults (e.g, memory bit-flips) and active attacks while designing privacy-preserving solutions for genome privacy. Therefore, the working group called for novel middleware genome data processing solutions that achieve scalability, availability, and performance along with integrity and confidentiality requirements.

The group discussed distributed computing and cryptographic approaches to overcome these challenges. A theme that emerged out of the discussion was to do data processing in a secure manner in a distributed/outsourced environment rather than doing it in an insecure and error-prone manner on the machines themselves locally. The group also discussed about designing distributed encrypted file systems and indexing solutions for large-scale genome data. From the cryptographic point of view, necessity of authenticated data structures and verifiable computations for genomic data was considered. The group briefly explored the possibility of using known cryptographic tools such as Merkle hash trees, watermarking, erasure coding, and secret sharing, and then called for tailored integrity protection and provenance solution for genomic data.

4.3 Inference Control

Muhammad Naveed (University of Illinois – Urbana-Champaign, US)

Genomic data can be used to infer identity, disease, traits, and kinship. Inference is not a static process and change over time; therefore, inference control procedures should similarly change over time. Management of inference is a complex process, and common people may not understand the all the implications. A neutral entity should provide guidance to the users to decide whether they should donate their data and the type of consent they should give. While inference risks can be reduced, it affects utility. Cryptographic methods cost more and statistical measures add noise. Communicating just the risks to the participants without discussing the potential benefits would be a disservice to the humanity. We have to balance the critical tradeoff between risk and utility, for example, U.S. federal agencies use RU-confidentiality curve to determine risk vs. utility tradeoffs.

We discussed what type of information would be a concern if it could be inferred from genomic data. Identity is a major concern, for example, if an adversary can infer whether an individual is in the case or control group of a genome-wide association study (GWAS). Attribute disclosure was also discussed and whether it constitutes a privacy breach. As attribute disclosure reveals information about population and no individual-level information, considering it a privacy breach could be detrimental to science. Incidental findings are a growing concern; that is, if a medical professional learns something about the patient that she did not ask for, whether the patient should be told or not. Physicians are also afraid of being sued if they fail to properly adjust dosage based on genomic information.

Genomics is still in its infancy and rapidly growing; therefore, we do not know what information or utility we require from genomic data. This makes addressing privacy concern even more challenging. We discussed if a general and evolving framework for specifying an inference of private data would be a reasonable approach. European Union agrees that anonymous data is a myth, but what type of data would they still be comfortable with releasing; after all, sharing of medical and genomic data is crucial for the development of medicine and technologies for human health. One reasonable approach could be that data owner (or data custodian) can ask people, whose data are being used, for what type of purposes their data could be used and what type of information could be inferred from their data. Another approach would be to specify necessary conditions that should be satisfied before one can obtain data for a study, for example, keeping detailed record of who uses the data and for what purpose. Such conditions could not be satisfied for public use summary statistics.

Sharing of genomic data is crucial for our understanding of disease and human health. The participants raised several points about sharing of genomic data, such as, can we share single nucleotide polymorphism (SNP) without knowing knowledge and power of the adversary? Or can we publish rare variants (e.g., occurring in a single person) in public datasets? Risk assessment is important here. Commonly used risk assessment models are geared towards worst-case adversaries. A challenge here is to develop reasonable risk models against plausible adversaries. The first step in addressing this challenge is to determine what risk is acceptable. For example, explaining the risk during the consent process such that the participants understand the actual risk, developing a comprehensive set of precautions and procedures data collectors should use, and respecting the right not to know. It looks like that research community does not like inference control, but maybe it makes sense in other contexts such as direct to consumer or legal contexts?

4.4 Usage models of genomic privacy

Kurt Rohloff (NJIT – Newark, US)

The goal of the working group was to identify the stakeholders for genomic applications. We commenced our analysis by identifying that storage and processing of genomic information could occur at multiple levels. These stakeholders include data producers, who turn samples into data, and sample owners, such as patients and cryobanks. The stakeholders also relate the concept of a "genome owner" who could either be a sample owner, or a designated representative, such as a physician.

The physician (or pathologist) is primarily responsible for consent. Once consent is granted, the sequencing facility receives the sample and performs sequencing. Interpretation is performed by a interpreter / bioninformatician. If there is a clinical motivation for sequencing, then clinical geneticist could be engaged. Results could also be sent to researchers, such as for population/public health studies. Results turned into text for the doctor about patient and actionable information. Relevant attacks on this ownership chain include insider attacks, such as by a network administrator. Data integrity is also perceived as an issue, either from adversaries, or from simple computing over large datasets where bit flips could happen that would alter data.

A feasible research scenario involves multiple data controllers contributing data, such as for collaboration projects which includes phenotypic data. A general question Is one of who has data with property X. Requests made in this scenario include one of identifying who has data in a specific format. A general approach, to improve performance, is to move the algorithm to the data.

Another challenge is dealing with the scenario of when consent is revoked. Participants need to be able to indicate to remove the data.

As an indication of scale, a participant would need to keep up to 4 million variants for a specific individual over a set of 200 million known variants. Every single 4 or 5 million.

Participants

- Luk Arbuckle
CHEO Research Institute – Ottawa, CA
- Erman Ayday
Bilkent University – Ankara, TR
- Marina Blanton
University of Notre Dame, US
- Dan Bogdanov
Cybernetica AS – Tartu, EE
- Emiliano De Cristofaro
University College London, GB
- Zekeriya Erkin
TU Delft, NL
- Jacques Fellay
EPFL – Lausanne, CH
- Kay Hamacher
TU Darmstadt, DE
- Zhicong Huang
EPFL – Lausanne, CH
- Jean Pierre Hubaux
EPFL – Lausanne, CH
- Mathias Humbert
Universität des Saarlandes, DE
- Aniket Kate
Universität des Saarlandes, DE
- Stefan Katzenbeisser
TU Darmstadt, DE
- Florian Kerschbaum
SAP AG – Karlsruhe, DE
- Oliver Kohlbacher
Universität Tübingen, DE
- Florian Kohlmayer
TU München – Klinikum Rechts der Isar, DE
- Alexander Kaitai Liang
Aalto University, FI
- Huang Lin
EPFL – Lausanne, CH
- Bradley Malin
Vanderbilt Univ. – Nashville, US
- Adam Molyneaux
Sophia Genetics SA – Lausanne, CH
- Muhammad Naveed
University of Illinois – Urbana-Champaign, US
- Jun Pang
University of Luxembourg, LU
- Fabian Prasser
TU München – Klinikum Rechts der Isar, DE
- Manuel Prinz
DKFZ – Heidelberg, DE
- Jean-Louis Raisaro
EPFL – Lausanne, CH
- Kurt Rohloff
NJIT – Newark, US
- Dominique Schröder
Universität des Saarlandes, DE
- Vitaly Shmatikov
University of Texas – Austin, US
- Sean Simmons
MIT – Cambridge, US
- Adam Davison Smith
Pennsylvania State University – University Park, US
- Thorsten Strufe
TU Dresden, DE
- Qiang Tang
University of Luxembourg, LU
- Carmela Troncoso
IMDEA Software – Madrid, ES
- Juan Ramon Troncoso Pastoriza
University of Vigo, ES
- Gene Tsudik
Univ. of California – Irvine, US
- Paulo Jorge Veríssimo
University of Luxembourg, LU
- Xiaofeng Wang
Indiana University – Bloomington, US

Report of Dagstuhl Seminar 15441

Duality in Computer Science

Edited by

Mai Gehrke[1], Achim Jung[2], Victor Selivanov[3], and Dieter Spreen[4]

1 **LIAFA, CNRS and Univ. Paris-Diderot, FR,** mgehrke@liafa.univ-paris-diderot.fr
2 **University of Birmingham, GB,** a.jung@cs.bham.ac.uk
3 **A. P. Ershov Institute – Novosibirsk, RU,** vseliv@iis.nsk.su
4 **Universität Siegen, DE,** spreen@math.uni-siegen.de

Abstract

This report documents the programme and outcomes of Dagstuhl Seminar 15441 'Duality in Computer Science'. This seminar served as a follow-up seminar to the seminar 'Duality in Computer Science' (Dagstuhl Seminar 13311). In this seminar, we focused on applications of duality to semantics for probability in computation, to algebra and coalgebra, and on applications in complexity theory. A key objective of this seminar was to bring together researchers from these communities within computer science as well as from mathematics with the goal of uncovering commonalities, forging new collaborations, and sharing tools and techniques between areas based on their common use of topological methods and duality.

Seminar October 25–30, 2015 – http://www.dagstuhl.de/15441

1998 ACM Subject Classification F.1.1 Models of Computation, F.3.2 Semantics of Programming Languages, F.4.1 Mathematical Logic, F.4.3 Formal Languages

Keywords and phrases coalgebra, domain theory, probabilistic systems, recognizability, semantics of non-classical logics, Stone duality

Digital Object Identifier 10.4230/DagRep.5.10.66

Edited in cooperation with Samuel J. van Gool

1 Executive Summary

Mai Gehrke
Achim Jung
Victor Selivanov
Dieter Spreen

Aims of the seminar

Duality allows one to move between an algebraic world of properties and a spacial world of individuals and their dynamics, thereby leading to a change of perspective that may, and often does, lead to new insights. Because computer science is fundamentally concerned both with specification of programs and the dynamics of their executions, dualities have given rise to active research in a number of areas of theoretical computer science. In this seminar we particularly wanted to concentrate on applications of duality in semantics for continuous data with special focus on probability in computation, algebra and coalgebra, and applications in complexity theory.

The seminar

Our call for participation was exceptionally successful and right up to the actual start of the meeting we were in danger of exceeding the number of places allocated. We see this as a vindication of our aim of bringing these researchers together for exchanging ideas centred around the common topic of duality. The talks offered fell quite naturally into groupings which allowed us to adopt a fairly thematic programme structure:

Day 1, morning session: Duality and classical algebra. Talks by Libor Barto, Michael Pinsker, Max Dickmann, and Marcus Tressl.

Day 1, afternoon session: Duality and categories. Talks by Paul Taylor, Steve Vickers, and Pino Rosolini.

Day 2, morning session: Duality and topology. Talks by Matthew de Brecht, Mathias Schröder, Reinhold Heckmann, and Jean Goubault-Larrecq.

Day 2, afternoon session: Alternative views on duality. Talks by Niels Schwartz, George Hansoul, Rob Myers, and Alexander Kurz.

Day 3, morning session: Duality and coalgebra. Talks by Adriana Balan, Dirk Pattinson, Ulrich Berger, and Samuel J. van Gool.

Day 4, morning session: Duality and domain theory. Talks by Jimmie Lawson, Abbas Edalat, Achim Jung, and Klaus Keimel.

Day 4, afternoon session: Duality and logic. Talks by Peter Schuster, Martín Escardó, Vladimir Shavrukov, and Vasco Brattka.

Day 5, morning session: Duality and probability. Talks by Willem Fouché, Dexter Kozen, Daniela Petrişan, and Drew Moshier.

Final thoughts

As always, Dagstuhl staff were incredibly efficient and helpful which allowed all of us, including the organisers, to focus on the exchange of ideas and plans for joint work. We are sincerely grateful to them for their hospitality and professionalism.

Mai Gehrke (LIAFA, CNRS and University Paris Diderot)
Achim Jung (School of Computer Science, University of Birmingham)
Victor Selivanov (Institute of Informatics Systems, RAS, Novosibirsk)
Dieter Spreen (Math. Logik und Theoretische Informatik, Universität Siegen)

2 Table of Contents

3 Overview of Talks

3.1 Some aspects of positive coalgebraic logic

Adriana Balan (University Politehnica of Bucharest, RO), Alexander Kurz (University of Leicester, GB), and Jiří Velebil

Main reference A. Balan, A. Kurz, J. Velebil, "Positive fragments of coalgebraic logics," Logical Methods in Computer Science, 11(3):1–51, 2015.
URL http://dx.doi.org/10.2168/LMCS-11(3:18)2015
URL http://arxiv.org/pdf/1402.5922

Positive modal logic was introduced in an influential 1995 paper of Dunn [2] as the positive fragment of standard modal logic, using modal formulas built only from atomic propositions, conjunction, disjunction, box and diamond. The corresponding semantics is based on Kripke frames, equivalently, on coalgebras for the powerset functor. In the present talk we will show how to generalize Dunn's result from Kripke frames to coalgebras for a (weak-pullback preserving) functor on the category of sets, using two well-known dual adjunctions: between sets and Boolean algebras, respectively between posets and bounded distributive lattices. To ensure that the resulting positive coalgebraic logic will indeed extend the logic of distributive lattices by monotone modal operations, we have to work in an ordered-enriched context.

This is joint work with A. Kurz and J. Velebil [1].

References

1 A. Balan, A. Kurz, and J. Velebil. Positive Fragments of Coalgebraic Logics. *Log. Methods Comput. Sci.*, 11(3:18):1–51, 2015.
2 J. M. Dunn. Positive Modal Logic. *Studia Logica*, 55(2):301–317, 1995.

3.2 The CSP Basics Revisited

Libor Barto (Charles University – Prague, CZ)

Joint work of Barto, Libor; Opršal, Jakub; Pinsker, Michal
Main reference L. Barto, J. Opršal, M. Pinsker, "The wonderland of reflections", Unpublished manuscript, Oct. 2015.
URL http://www.karlin.mff.cuni.cz/~barto/Articles/0_doubleshrink.pdf

There are three reductions used to compare the complexity of constraint satisfaction problems (CSPs) over two relational structures: a) reduction by means of pp-interpretation, b) via homomorphic equivalence, c) by adding constants to cores. A fundamental fact for the CSP theory is that pp-interpretations between relational structures correspond to continuous homomoprhisms between the algebraic alter egos of these structures – their polymorphism clones. We extend this fact to include the other two reductions.

As a corollary, we get that the complexity of the CSP over a finite structure X depends only on height 1 identities satisfied by polymorphisms of X. The role of continuity for CSPs over infinite relational structures will be discussed as well.

3.3 Duality and Concurrency in Program Extraction

Ulrich Berger (Swansea University, GB)

Program extraction from proofs combines logic and programming to a method for creating provably correct software [1]. In this talk we look at two important principles, one from logic and one from programming, that so far were out of reach for program extraction, and propose ways to change this.

The principle in question from logic is duality in categories, more specifically, the duality of initial algebras and terminal coalgebras which correspond to induction and coinduction. In constructive logic or type theory (the usual formalisms behind program extraction) this duality seems broken, and while induction are constructively well-understood the constructive theory of coinduction is still in a state of flux.

The programming principle in question is concurrency, that is, the possibility that computations are executing simultaneously, are potentially interacting with each other, and may deliver conflicting results. In addition, some of the simultaneous computations may diverge. In traditional program extraction all computations terminate and there is no concurrency.

We will introduce a formal system for program extraction that reinstalls the duality between induction and coinduction and allows concurrency. The latter will be illustrated by applying program extraction to Tsuiki's infinite Gray-code [2].

References

1 U. Berger, K. Miyamoto, H. Schwichtenberg, and M. Seisenberger. Minlog – a tool for program extraction for supporting algebra and coalgebra. In *CALCO-Tools*, volume 6859 of *LNCS*, pages 393–399. Springer, 2011.

2 H. Tsuiki. Real Number Computation through Gray Code Embedding. *Theor. Comput. Sci.*, 284(2):467–485, 2002.

3.4 Residuals in the Weihrauch lattice and their applications

Vasco Brattka (Universität der Bundeswehr – München, DE)

Joint work of Vasco Brattka with Arno Pauly and Matthew Hendtlass, Alexander P. Kreuzer
Main reference V. Brattka, M. Hendtlass, A. P. Kreuzer, "On the Uniform Computational Content of Computability Theory," arXiv:1501.00433v2 [math.LO], 2015.
URL http://arxiv.org/abs/1501.00433v2

We discuss the algebraic structure of the Weihrauch lattice. Besides infimum and supremum, the lattice comes equipped with a product and a compositional product as well as two natural closure operators, the star product and the parallelization. Hence, the Weihrauch lattice carries a rich algebraic structure and with a certain subset of these operations it is, for instance, a Kleene algebra. In this talk we address the fact that none of the operations inf, sup and product have left or right residuals, but the compositional product $*$ has a one-sided residual that we call implication $\rightarrow$. This residual has a number of interesting applications and can be used to characterize computational problems such as Martin-Löf randomness MLR or cohesiveness COH. For instance, we obtain $\mathrm{MLR} \equiv_{\mathrm{W}} (\mathrm{C}_{\mathbb{N}} \rightarrow \mathrm{WWKL})$ and $\mathrm{COH} \equiv_{\mathrm{W}} (\lim \rightarrow \mathrm{KL})$, where $\mathrm{C}_{\mathbb{N}}$ denotes choice on the Natural numbers, KL denotes König's Lemma and WWKL Weak Weak König's Lemma.

3.5 Duality theory for quasi-Polish and represented spaces

Matthew de Brecht (NICT – Osaka, JP)

We will introduce some recent duality-related research within three categories of increasing generality.

The first topic concerns the category of quasi-Polish spaces. An important result by R. Heckmann shows that these spaces dually correspond to countably presented frames. Stated differently, quasi-Polish spaces can be viewed as the spaces of models of countably axiomatized propositional geometric theories. We will introduce joint work with T. Kawai showing that this correspondance naturally extends to the standard powerspace / powerlocale constructions. This provides a connection between quasi-Polish co-algebras and propositional geometric modal logic.

The second topic concerns the (cartesian closed) category of QCB-spaces. Recent work by M. Schröder and V. Selivanov organized these spaces into a hierarchy according to the complexity of their admissible representations. A separate but related hierarchy was later defined by Schroder, Selivanov, and the current author which characterizes spaces according to the complexity of defining a basis for their topology. Based on the hierarchies, we will raise the question of whether the duality characterizations of quasi-Polish spaces have natural generalizations within the category of QCB-spaces.

The third topic concerns the (locally cartesian closed) category of (Baire-) represented spaces. We will introduce some joint work with A. Pauly which uses Sierpinski-like objects to classify Borel subsets of spaces. We show that some of the basic results of P. Taylor's Abstract Stone Duality have natural interpretations in this setting. In particular, we provide an "abstract" version of the Jayne-Rogers theorem, which is of interest in descriptive set theory.

This work was supported by JSPS Core-to-Core Program, A. Advanced Research Networks and by JSPS KAKENHI Grant Number 15K15940.

3.6 Exploring the domain-theoretic content of some mathematical structures

Max Dickmann (University of Paris VII, FR)

The first half of the talk was devoted to:
1. Indicate the relationship between the topology of a spectral space and the Scott topology of its specialization order.
2. Characterize in terms of order the spectrality of the Scott topology on forests and root systems, and determine its relation with the coarse lower topology in these and other, related cases.

In the second half we stated some outstanding domain-theoretic properties of the following structures:
- The spaces of preorders and of quadratic modules of (commutative, unitary) rings where -1 is not a sum of squares.

- Various classes of prime ideals of the ring of continuous, real-valued functions on a topological space.
- Root systems of finite Krull dimension.

The results reported in this talk will appear, among many other related results, in Chapter 11 of the forthcoming book [1].

References

1 M. Dickmann, N. Schwartz, M. Tressl. *Spectral Spaces*. Cambridge University Press (approx. 450 pages; to appear in 2017).

3.7 Differentiation in Logical Form

Abbas Edalat (Imperial College London, GB)

Joint work of Mehrdad Maleki
Main reference A. Edalat, "A continuous derivative for real-valued functions," in S. B. Cooper, B. Lowe, A. Sorbi (eds.), "New Computational Paradigms, Changing Conceptions of What is Computable," LNCS, Vol. 4497, pp. 493–519, Springer, 2008; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/978-3-540-73001-9_26
URL http://www.doc.ic.ac.uk/~ae/papers/banachff.pdf
Main reference A. Edalat, A. Lieutier, D. Pattinson, "A Computational Model for Multi-Variable Differential Calculus," Information and Computation, Vol. 224, pp. 22–45, 2013.
URL http://dx.doi.org/10.1016/j.ic.2012.11.006

We introduce a point free, localic representation of the Clarke gradient of a real valued locally Lipschitz map on a finite dimensional Euclidean space. Our work is based on the domain-theoretic derivation of the Clarke gradient, called the L-derivative, which defines the Clarke gradient of a Lipschitz map as the supremum of the single-step functions that correspond to the single-ties the function satisfies. Here, a single-tie is a collection of locally Lipschitz maps that have a generalised compact, convex set-valued Lipschitz constant in an open neighbourhood in the Euclidean space. We introduce the notion of a strong single-tie which has the following property: The L-derivative of a Lipschitz map is the supremum of single-step functions, each way-below the L-derivative, that correspond to the strong single-ties the function satisfies. For the localic representation of the L-derivative using approximable mappings, we define a strong knot of approximable mappings, with respect to a given open neighbourhood and and an open set, which satisfies the following stone duality: If a Lipschitz map is in the strong tie of an open neighbourhood and a compact convex set, then the approximable mapping representing the Lipschitz map is in the knot of the open neighbourhood and any open set containing the compact convex set. Conversely, if an approximable mapping is in the strong knot of an open neighbourhood and an open set then there exists a compact convex subset of the open set such that the Lipschitz map corresponding to the approximable mapping is in the strong tie of the open neighbourhood and any open set containing the compact convex set. We then show that this duality enables us to obtain the L-derivative in localic framework.

3.8 Compactly generated Hausdorff locales

Martin H. Escardo (University of Birmingham, GB)

We say that a Hausdorff locale is compactly generated if it is the colimit of the diagram of its compact sublocales connected by inclusions. We show that this is the case if and only if the natural map of its frame of opens into the second Lawson dual is an isomorphism. More generally, for any Hausdorff locale, the second dual of the frame of opens gives the frame of opens of the colimit. In order to arrive at this conclusion, we generalize the Hofmann–Mislove–Johnstone theorem and some results regarding the patch construction for stably locally compact locales. In particular, the colimit is the patch of the first Lawson dual of the frame.

3.9 Topological dualities in structural Ramsey theory

Willem L. Fouché (UNISA – Pretoria, ZA)

Many results in classical Ramsey theory and the theory of well quasi-orderings can be proven by topological means, or be expressed as topological-dynamical phenomena.

Thierry Coquand has given many examples of such topological (-dynamical) expressions of combinatorial phenomena (Kruskal a la Nash Williams, the countable version of van der Waerden), which are classically equivalent to statements involving well-quasi-orderings in point-free topology, or locales, and which are, thus formulated, constructively provable.

I studied topological versions of these combinatorial phenomena from the viewpoint of Gelfand duality of commutative C^*-algebras, the latter duality being provable in constructive mathematics, when adequately phrased, and having, therefore, interesting computational content.

References

1 Andreas Blass. Prime ideals yield almost maximal ideals. *Fund. Math*, 127:57–66, 1987.
2 W. L. Fouché. Symmetry and Ramsey degrees of finite relational structures. *J. Comb. Theory A*, 85:135–147, 1997.
3 P. Freyd All topoi are localic, or Why permutation models prevail. *J. Pure Appl. Alg.*, 46:49–58, 1987.
4 A. S. Kechris, V. G. Pestov, and S. Todorcevic. Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups. *Geometric And Functional Analysis*, 15:106–189, 2005.

3.10 Pointfree convergence

Jean Goubault-Larrecq (ENS – Cachan, FR)

Joint work of Jean Goubault-Larrecq and Frédéric Mynard

The objective of this work is to define a pointfree analogue of convergence spaces, mirroring the theory of locales as a pointfree analogue of topological spaces. The latter rests on a famous duality, which evolved over the ages from the work of Marshall Stone: there is a functor $\mathcal{O}_{\text{Stone}}$ from the category **Top** of topological spaces to the category of locales, that is, to the opposite of the category of frames, and $\mathcal{O}_{\text{Stone}}$ is left adjoint to a functor $\mathsf{pt}_{\text{Stone}}$. This reduces to a duality between sober spaces and spatial lattices.

Convergence spaces, due to G. Choquet, form a strictly larger category than **Top**, and one that has some features that **Top** lacks, for example Cartesian closure. We build a similar adjunction between the category **Conv** of convergence spaces and a category of so-called *focales*, study its associated monad, and relate this adjunction to the Stone adjunction mentioned above.

Precisely, we define a convergence lattice as a bounded lattice L together with a monotone map lim from the poset of filters on L to L. Every convergence space X gives rise to a convergence lattice $\mathcal{L}(X)$, obtained as its powerset. This defines a functor $\mathcal{L}$ from the category of convergence spaces to the opposite of the category of convergence lattices, and that is what we call the category of focales.

We show that this functor has a right adjoint pt. For a convergence lattice L, $\mathsf{pt}L$ is the convergence space of all compact prime filters of L, with some suitable notion of convergence.

Any adjunction gives rise to a monad. In the case of the Stone adjunction, the result monad is the sobrification monad. We study the monad associated with our $\mathcal{L} \dashv \mathsf{pt}$ adjunction, yielding two possible notions of sobriety for convergence spaces:

- the *temperate* convergence spaces X are those that can be written as the image of some convergence lattice L by the monad functor, and we characterize them as those spaces that are replete, tiled, and separated with respect to some so-called designated limit function; we can take for L the convergence lattice of all *tiles* on X—tiles are a fundamental notion to our new theory;
- the *tee-totalers* are those such that the unit of the monad is an isomorphism; we characterize them as those convergence spaces where every compact ultrafilter is principal, or equivalently where every subset is a tile.

Every tee-totaler is temperate, and every temperate convergence space is quasi-sober, where quasi-sobriety means that every compact ultrafilter has a generic point. All three notions are equivalent for T_0 topological spaces.

Reinhold Heckmann presented another list of possible notions of sobriety for Ω-embedded convergence spaces at this Dagstuhl workshop. There are striking similarities in some of our definitions. We have started to compare our notions, but haven't yet reached a conclusion.

In a second part, we study the relationship between our adjunction $\mathcal{L} \dashv \mathsf{pt}$ and the familiar Stone adjunction between topological spaces and locales. The two adjunctions fit nicely into a commuting diagram of four adjunctions, as we have shown. The last open problem that remained was solved by the audience at Dagstuhl.

3.11 Mereotopological distributive lattices

Georges Hansoul (University of Liège, BE)

Mereotopological structures have been devised to algebraize Whitehead's conception of space and time (by regions and intervals rather than points and instants). They also serve as models for the RCC (region connection calculus) and some logics such as MTML. These are structures over a signature containing mereological (i.e. boolean) symbols and symbols of a more topological nature ($a \prec b$: a is a non-tangential part of b: $\bar{a} \subseteq b^{\circ}$; $a\mathcal{C}b$: a is in contact with b: $\bar{a} \cap \bar{b} \neq 0$; ...). Among these structures, well known are contact algebras, for which we give two definitions. Let B be a boolean algebra. If $\mathcal{C} \subseteq B^2$, then $\mathcal{C}(a,-)$ denotes $\{b \mid a\mathcal{C}b\}$ and $\not\mathcal{C}$ is $B^2 \setminus \mathcal{C}$.

1. A structure $\mathcal{B} = (B, \mathcal{C})$ is a *contact algebra* if $\mathcal{C} \subseteq B^2$ satisfies:
   (i) $\not\mathcal{C}(a,-)$ and $\not\mathcal{C}(-,a)$ are ideals for each $a \in B$;
   (ii) $a\mathcal{C}a \to a = 0$;
   (iii) $\mathcal{C}$ is symmetric;
   (iv) $a \mapsto \mathcal{C}(a,-)$ is one-to-one.
5. A structure $\mathcal{B} = (B, \prec)$ is a *proximity algebra* if $\prec \subseteq B^2$ satisfies:
   (i) $\prec (a,-)$ is a filter and $\prec (-,a)$ an ideal for each $a \in B$;
   (ii) $a \prec b \to a \leq b$;
   (iii) $a \prec b \to \neg a \prec \neg b$ ($\neg$ is for complement);
   (iv) $a = \vee\{b \mid b \prec a\}$.

These definitions are equivalent through the link: $(*) \quad a \prec b \quad \text{iff} \quad (a \not\mathcal{C} \neg b)$.

And one of the most fruitful results in the theory is a representation theorem which states ([1]) that any contact algebra may be densely embedded in a *standard* contact algera, that is a contact algebra $(RO(Y), \mathcal{C}_\tau)$ where $Y = (Y, \tau)$ is a topological space, $RO(Y)$ is the boolean algebra of all regular open subsets of Y and, for $O, V \in RO(Y)$, $O\mathcal{C}_\tau V$ iff $\bar{O} \cap \bar{V} \neq \emptyset$.

This result is in fact very closed to a result obtained long before by de Vries for completely different purposes and celebrating a duality between compact spaces and de Vries algebras (complete proximity algebras satisfying $a \prec b \to \exists\, c,\ a \prec c \prec b$).

In [2] was raised the need to extend the theory to structures where the mereological part is no longer boolean and $\prec$, $\mathcal{C}$ are no longer interdefinable. They give some representation theorem for contact distributive lattices $(L, \mathcal{C})$. On the other hand, Bezhanishvili and Harding extend de Vries duality to *proximity frames* $(L, \prec)$. But what is really asked in [2] is a study for mixed structures $(L, \mathcal{C}, \prec)$. Two problems are examined: 1) when can we say that $\mathcal{C}$ and $\prec$ are linked (that is, convey to a unique conception of time or space) and 2) do we have representation theorems?

A *mereotopological lattice* is a structure $\mathcal{L} = (L, \mathcal{C}, \prec)$ where L is a bounded distributive lattice and $\mathcal{C}$, $\prec$ satisfies their respective axiom (i).

▶ **Theorem 1.** *For a mereotopological lattice $\mathcal{L}$, the following are equivalent:*
1. *for each $a \in L$, $\prec (a,-)$ and $\not\mathcal{C}(a,-)$ are kernel and cokernel of some congruence on L;*
2. (a) *$a \prec c$ and $a\mathcal{C}b \to a\mathcal{C}(b \wedge c)$ and* (b) *$a \prec b \vee c \to a \prec c$ or $a\mathcal{C}b$.*

▶ **Theorem 2.** *Let $\mathcal{L}$ be a mereotopological lattice satisfying $\prec$(ii), $\prec$(iv), $a\mathcal{C}b$ iff $\exists\, c$, $a \prec c$ and $b \wedge c = 0$, and $\mathcal{C}$ is symmetric.*

Then there exists a bitopological space $Y = (Y, \tau_1, \tau_2)$ and a dense embedding $\mathcal{L} \to RO_{12}(Y)$ where $O \in RO_{12}(Y)$ if $O = \mathrm{Int}_{\tau_2}\, cl_{\tau_1}(O)$, $O \prec U$ iff $cl_{\tau_1}(O) \subseteq U$ and $O\mathcal{C}U$ iff $cl_{\tau_1}(O) \cap cl_{\tau_1}(U) \neq \emptyset$.

References

1 Düntsch I., Winter M., A representation theorem for Boolean contact algebras, Theoret. Comput. Sci. (B) 347 (2005) 498–512.
2 Düntsch I., MacCaull W., Vakarelov D., Winter M., Distributive contact lattices: Topological representations, The Journal of Logic and Algebraic Programming 76 (2008), 18–34.

3.12 Notions of Sobriety for Convergence Spaces

Reinhold Heckmann (AbsInt – Saarbrücken, DE)

In this talk, we discuss and compare two notions of sobriety for convergence spaces, and relate them to repleteness. The first ("E-sober") has a very general definition in terms of a regular subspace of the double exponential monad w.r.t. Sierpinski space (as considered by Rosolini for equilogical spaces and Taylor in ASD), but has the drawback that almost no closure properties can be shown for the class of E-sober spaces. The second ("A-sober") has a quite special definition based directly on filters and convergence, but has the advantage that the class of "A-sober" spaces is closed under product, regular subspace, and exponentiation. Finally, we compare the two notions of sobriety with repleteness. Every E-sober space is replete and every replete space is A-sober. For topological spaces, both notions of sobriety coincide with the traditional sobriety, and so, topological spaces are replete in the category of convergence spaces if and only if they are sober.

3.13 On the Ho-Zhao problem

Achim Jung (University of Birmingham, GB)

Joint work of Achim Jung, Weng Kin Ho, and Xiaoyong Xi

Given a poset P, the set $\Gamma(P)$ of all Scott closed sets ordered by inclusion forms a complete lattice which is order-dual to the frame of open sets. A subcategory $\mathbf{C}$ of $\mathbf{Pos}_d$ (the category of posets and Scott-continuous maps) is said to be Γ-faithful if for any posets P and Q in $\mathbf{C}$, $\Gamma(P) \cong \Gamma(P)$ implies $P \cong Q$. It is known that the category of all continuous dcpo's and the category of bounded complete dcpo's are Γ-faithful, while $\mathbf{Pos}_d$ is not. Ho & Zhao (2009) asked whether the category **DCPO** of dcpo's is Γ-faithful. In this talk I presented an example that shows that this is not the case, thus settling the Ho-Zhao problem in the negative.

References

1 W. K. Ho and D. Zhao. Lattices of Scott-closed sets. *Comment. Math. Univ. Carolinae*, 50(2):297–314, 2009.

3.14 The Cuntz semigroup for C*-algebras and domain theory

Klaus Keimel (TU Darmstadt, DE)

Domain theory has its origins in Mathematics and Theoretical Computer Science. Mathematically it combines order and topology. Its central concepts have their origin in the idea of approximating ideal objects by their relatively finite or, more generally, relatively compact parts.

The development of domain theory in recent years was mainly motivated by question in denotational semantics and the theory of computation. But since in 2008 Coward, Elliott and Ivannescu have introduced a new invariant for C^*-algebras, domain theoretical notions and methods are used in the theory of C^*-algebras in connection with the Cuntz semigroup.

The talk was largely expository. It presented those notions of domain theory that seem to be relevant for the theory of Cuntz semigroups and have sometimes been developed independently in both communities. It also contains a new aspect in presenting results of Elliott, Robert and Santiago on the cone of traces of a C^*-algebra as a particular case of the dual of an abstract Cuntz semigroup.

3.15 Kolmogorov Extension, Martingale Convergence, and Compositionality of Processes

Dexter Kozen (Cornell University, US)

We show that the Kolmogorov extension theorem and the Doob martingale convergence theorem are two aspects of a common generalization, namely a colimit-like construction in a category of Radon spaces and reversible Markov kernels. The construction provides a compositional denotational semantics for standard iteration operators in probabilistic programming languages, e.g. Kleene star or while loops, as a limit of finite approximants, even in the absence of a natural partial order.

3.16 Stone Duality for Categories of Relations

Alexander Kurz (University of Leicester, GB)

Joint work of Achim Jung, Alexander Kurz, and Andrew Moshier

We show that the dual equivalence between proximity lattices and compact ordered Hausdorff spaces can be recovered by three well-known categorical constructions. First, the duality of distributive lattices and Priestley spaces is extended to weakening-closed relations. Second, idempotents are split. Third, maps are recovered as adjoint pairs of relations. Adjoint pairs homming into 2 are Dedekind cuts, thus giving an abstract account of extending duality from the zero-dimensional to the Hausdorff situation. The three steps can be described in purely category theoretic terms and can be applied to a range of different dualities.

3.17 Pointwise Directed Families and Continuity of Function Spaces

Jimmie D. Lawson (Louisiana State University – Baton Rouge, US)

(Up-)directed sets play an important role in domain theory and elsewhere, e.g., lower semicontinuous functions, idempotent analysis. In the study of function spaces $[X \to P]$, P a poset, X. Xi and J. Liang in 2009 introduced a weaker notion of directedness for functions, namely a family of functions F such that for each x in X, the set $\{f(x) \mid f \in F\}$ is a directed subset. Such a family of functions is called pointwise directed. Basic properties of pointwise directed families will be presented; the principal application is showing that certain new examples of classes of function spaces are continuous domains.

3.18 Natural Duality for Representable Generalized Orthoalgebras

M. Andrew Moshier (Chapman University – Orange, US)

Joint work of M. Andrew Moshier and P. Jipsen

Separation Logic is meant to formalize allocation of resources (memory, processors, and so on). A resource may be allocated or not and cannot be allocated twice. Heaps constitute a typical class of structures that can be modelled this way, where a heap is the collection of all partial functions from a set L (locations) to V (values). In a heap, $\bot$ is the entirely undefined function and $\oplus$ is merging of partial functions ony when they have disjoint domains. Heaps, as well as other sorts of resources, constitute partial cancellative, positive commutative monoid (in the literature, known as generealized orthoalgebras [GAOs]). We study this idea from the perspective of universal algebra and natural duality.

Consider the simplest posssible heap P_1 consisting of the partial functions from $\{1\}$ to $\{1\}$. We can write this as $\{\bot, \top\}$ with $\bot$ the undefined function, $\top$ the defined function and the partial operation $\oplus$ for which $\top \oplus \top$ is undefined and $x \oplus y$ behaves like join otherwise. We consider the quasivariety $ISP(P_1)$. As a confirmation that this quasivariety captures something useful, we note that all heaps belong to $ISP(P_1)$. Of course, there are other structures in the quasivariety that are not simple heaps, but they are all representable as heaps with allocation constraints rouggly of the form "if $h(l_1)$ is allocated then $h(l_2)$ is also."

We show that $ISP(P_1)$ is not finitely axiomatizable by quasi-equations. We then show that in spite of this, $ISP(P_1)$ is naturally dualizable at the finite level. That is, there is a structure Q_1 on the same underlying set $\{\bot, \top\}$ so that $ISP_\omega(P_1)$ is dually equivalent to $ISP_\omega(Q_1)$ (P_ω meaning we take only finite products) via the usual construction of hom-sets into the dualizing object. As this is work in progress, we discuss in some detail the problems involved in extending the finite-level duality to a duality for all of $ISP(P_1)$.

3.19 Program Equivalence is Coinductive

Dirk Pattinson (Australian National University – Canberra, AU)

We describe computational models, notably Turing and counter machines, as state transition systems with side effects. Side effects are expressed via an algebraic signature and interpreted over comodels for that signature; basically, comodels describe the memory model while the transition system captures the control structure. Completeness of equational reasoning over comodels is known to be a subtle issue. We identify a criterion on equational theories and classes of comodels that guarantees completeness, over the given class of comodels, of the standard equational calculus, and show that this criterion is satisfied in our leading examples. Based on such a complete equational axiomatization of the memory model, we then give a complete inductive-coinductive calculus for program equivalence in the full computational model. This calculus is phrased in terms of simulation between states, where a state simulates another if it has at least the same terminating computations, with the same cumulative effect on global state.

3.20 Coinduction up-to techniques in a fibrational setting

Daniela Petrişan (University of Paris VII, FR)

Joint work of Filippo Bonchi, Daniela Petrişan, Damien Pous, and Jurriaan Rot

Bisimulation is used in concurrency theory as a proof method for establishing behavioural equivalence of processes. Up-to techniques can be seen as a means of optimizing proofs by coinduction. For example, to establish that two processes are equivalent one can exhibit a smaller relation, which is not a bisimulation, but rather a bisimulation up to a certain technique, say 'up-to contextual closure'. However, the up-to technique at issue has to be sound, in the sense that any bisimulation up-to should be included in a bisimulation.

In this talk, I will present a general coalgebraic framework for proving the soundness of a wide range of up-to techniques for coinductive unary predicates, as well as for bisimulations. The specific up-to techniques are obtained using liftings of functors to appropriate categories of relations or predicates. In the case of bisimulations with silent moves the situation is more complex. Even for simple examples like CCS, the weak transition system gives rise to a lax bialgebra, rather than a bialgebra. In order to prove that up-to context is a sound technique we have to account for this laxness. The flexibility and modularity of our approach, due in part to using a fibrational setting, pays off: I will show how to obtain such results by changing the base category to preorders.

This is joint work with Filippo Bonchi, Damien Pous and Jurriaan Rot and is based on the papers [1] and [2].

References

1 Filippo Bonchi, Daniela Petrişan, Damien Pous, and Jurriaan Rot. Coinduction up-to in a fibrational setting. In *Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on*

Logic in Computer Science (LICS), CSL-LICS'14, Vienna, Austria, 2014, pages 20:1–20:9. ACM, 2014.
2 Filippo Bonchi, Daniela Petrişan, Damien Pous, and Jurriaan Rot. Lax bialgebras and up-to techniques for weak bisimulations. In *26th International Conference on Concurrency Theory, CONCUR 2015, Madrid, Spain, 2015*, volume 42 of *LIPIcs*, pages 240–253. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015.

3.21 Uniform Birkhoff

Michael Pinsker (Charles University – Prague, CZ)

Joint work of Michael Pinsker and Mai Gehrke

I will present a joint work with Mai Gehrke which characterizes pseudovarieties of finitely generated algebras, i.e., classes of finitely generated algebras closed under homomorphic images, subalgebras, and finite products. The result involves both algebraic and topological considerations, and is a common generalization of a description of pseudovarietes of finite algebras due to Eilenberg, Schuetzenberger, Reiterman and Banaschewski, as well as the description of the pseudovariety generated by a single oligomorphic algebra due to Bodirsky and myself. I will also outline the connection of pseudovarieties with Constraint Satisfaction Problems.

3.22 Frames as equilogical algebras

Giuseppe Rosolini (University of Genova, IT)

Joint work of Giuseppe Rosolini, Giulia Frosoni, and Alessio Santamaria

We present a connection between frames and algebras for the double exponential monad on the Sierpinski space. I spoke about this in the last Dagstuhl seminar on Duality and we are now close to a complete solution of the problem. Instrumental for the presentation is Dana Scott's category of equilogical spaces, see [1, 2].

I also connect our work on algebras for the double exponential monad with the work of Paul Taylor on Abstract Stone Duality [3] and with that of Steve Vickers of algebras for the double powerlocale monad [4].

References

1 Andrej Bauer, Lars Birkedal, Dana S. Scott. *Equilogical spaces*. Theoret. Comput. Sci. 315 (2004)
2 Dana S. Scott. *Data types as lattices*. SIAM J. Comput. 5 (1976)
3 Paul Taylor. *Foundations for computable topology*. Foundational theories of classical and constructive mathematics, West. Ont. Ser. Phil. Sci. 76 (2011)
4 Steve Vickers. *The double powerlocale and exponentiation: a case study in geometric logic*. Th. Appl. Cat. 12 (2004)

3.23 Towards a theory of sequentially locally convex QCB-spaces

Matthias Schröder (TU Darmstadt, DE)

We develop a theory of sequentially locally convex QCB-spaces. These are vector spaces equipped with a QCB-topology that arises as the sequentialisation of a locally convex topology. Remember that QCB-spaces form the class of topological spaces which can be handled by the representation based approach to Computable Analysis; a locally convex space is a topological vector space such that its topology is induced by a family of seminorms.

We investigate the class of sequentially locally convex co-Polish spaces. Co-Polish spaces are Hausdorff spaces such that the Scott-topology on the lattice of opens is quasi-Polish. They admit an admissible representation with a locally compact domain. Spaces equipped with such a representation allow for a *Simple Complexity Theory*, meaning the measurement of time complexity in terms of (1) a *discrete* (rather than a continuous) parameter on the input and (2) the output precision.

We present a duality result between the class of sequentially locally convex co-Polish spaces and and the class of separable metrisable locally convex spaces.

3.24 Eliminating Disjunctions by Disjunction Elimination

Peter M. Schuster (University of Verona, IT)

Completeness theorems, or contrapositive forms of Zorn's Lemma, are often invoked in elementary contexts in which the corresponding conservation theorems would suffice. Unlike the former, the latter are syntactical a priori and thus predisposed to proofs with finite methods only. To get a universal constructive conservation theorem for definite Horn clauses, we use Scott's multi-conclusion entailment relations as extending Tarskian single-conclusion consequence relations, or algebraic closure operators. In a nutshell, the extra multi-conclusion axioms can be reduced to rules for the underlying single-conclusion relation that in many contexts turn out to hold. Thanks to a sandwich criterion due to Scott, the method can also be proved optimal.

Applications include the separation and extension theorems known under the names of Krull-Lindenbaum, Artin-Schreier, Szpilrajn and Hahn-Banach. Related work can be found, for example, in locale theory (Mulvey-Pelletier 1991), dynamical algebra (Coste-Lombardi-Roy 2001, Lombardi 1997-8), formal topology (Cederquist-Coquand-Negri 1998) and proof theory (Coquand-Negri-von Plato 2004, Negri-von Plato 2011).

3.25 Locales as spectral spaces

Niels Schwartz (Universität Passau, DE)

Recently a new presentation of locales has been introduced, cf. [4]. Traditionally, locales are the objects of $\mathtt{Fr}^{op}$, the opposite of the category of frames, [3]. Stone duality is an anti-equivalence between the category `BDLat` of bounded distributive lattices and the category `Spec` of spectral spaces, [2], [1]. Frames are bounded distributive lattices and frame homomorphisms are bounded lattice homomorphisms. Therefore Stone duality associates with every frame L its spectrum $\mathrm{Spec}(L)$ and with every frame homomorphism $\varphi : L \to M$ the spectral map $\mathrm{Spec}(\varphi) : \mathrm{Spec}(M) \to \mathrm{Spec}(L)$. The spectra of frames are called *locales* and the Stone duals of frame homomorphisms are called *localic maps*. The locales and the localic maps form a subcategory `Loc` ⊂ `Spec`, which is called the *category of locales*.

Frames are the main objects in the study of pointfree topology. The new approach to locales introduces points into pointfree topology and makes it possible to view pointfree topology as a part of classical point set topology.

The lecture gives a brief explanation of basic facts about topological locales from [4] and current research directions, cf. [5], [6] and other forthcoming papers.

Locales are special spectral spaces and localic maps are special spectral maps. The following results characterize locales and localic maps by a simple topological condition. The characterization uses the *patch topology* or *constructible topology* of a spectral space, [2], [1]. The patch space of a spectral space X is denoted by X_{con}, the patch closure of a subset $M \subseteq X$ is denoted by $\overline{M}^{\mathrm{con}}$.

▶ **Theorem 1.** *A spectral space X is a locale if and only if $\overline{U}^{\mathrm{con}}$ is open for every open set $U \subseteq X$.*

▶ **Theorem 2.** *A spectral map $f : X \to Y$ between locales is a localic map if and only if $\overline{f^{-1}(V)}^{\mathrm{con}} = f^{-1}(\overline{V}^{\mathrm{con}})$ for each open $V \subseteq Y$.*

These results show that the formation of the patch closure of an open set is of basic importance for locales. Let X be a locale, $\mathcal{O}(X)$ its frame of open sets and $\mathring{\mathcal{K}}(X)$ its lattice of quasi-compact open sets.

▶ **Theorem 3.** *For a locale X the map $U \mapsto \overline{U}^{\mathrm{con}}$ is a nucleus of the frame $\mathcal{O}(X)$. Its image is the frame $\mathring{\mathcal{K}}(X)$. (This nucleus is called the* natural nucleus *of $\mathcal{O}(X)$.)*

These results are the starting point of a topological theory of locales. By Stone duality the category `Loc` is equivalent to $\mathtt{Fr}^{op}$. So there are numerous questions about the translation of facts about locales into facts about the objects of $\mathtt{Fr}^{op}$. For example, recall that the nuclei of a frame L are considered as the *sublocales* of the locale L, [3]. In the topological theory of locales a subset A of a locale X is called a *localic subspace* if it is a spectral subspace (i.e., is closed for the patch topology, [1]), is a locale with the subspace topology and the inclusion map $A \to X$ is a localic map.

▶ **Theorem 4.** *Let X be a locale and $A \subseteq X$ a spectral subspace. Then the following conditions are equivalent.*

1. *A is a localic subspace.*
2. *If $U \subseteq A$ and $V \subseteq X$ are open and $U = V \cap A$ then $\overline{U}^{\mathrm{con}} = \overline{V}^{\mathrm{con}} \cap A$.*
3. *If $C \subseteq A$ is constructible then $\overline{C}^{X}$, the closure in X, is constructible.*

The following category theoretic facts describe the relation of the category `Loc` with the categories `Top` of topological spaces and `Spec` of spectral spaces.

▶ **Theorem 5.** *The subcategory* `Spec` $\subset$ `Top` *is reflective. If X is a topological space then the locale $L(X) = \mathrm{Spec}(\mathcal{O}(X))$ is the reflection of X. The reflection map $S_X : X \to L(X)$ sends $x \in X$ to the prime ideal $\{O \in \mathcal{O}(X) \mid x \notin O\}$.*

▶ **Theorem 6.** *The subcategroy* `Loc` $\subset$ `Spec` *is coreflective. The coreflection of a spectral space X is the locale $L(X) = \mathrm{Spec}(\mathcal{O}(X))$. The coreflection map $R_X : L(X) \to X$ is the Stone dual of the inclusion homomorphism $\overset{\circ}{\mathcal{K}}(X) \to \mathcal{O}(X)$.*

Following these results it is an obvious question how properties of a topological space X are connected with properties of the spectral reflection $L(X)$. As examples of this type of results we mention:

▶ **Theorem 7.** *Let X be a spectral map. Then the following conditions are equivalent.*
1. *The reflection map S_X is a homeomorphism.*
2. *The reflection map is a spectral map.*
3. *X is Noetherian.*

▶ **Theorem 8.** *A topological space X is quasi-compact if and only if the spectral reflection $L(X)$ is quasi-compact.*

The spectral reflection can be used to construct the Stone-Čech compactification of a normal topological space and the Gleason cover of a compact space. In the literature various topological spaces have been associated with a partially ordered set. The spectral reflection can be used to relate many of these constructions to each other and to produce new constructions.

Finally, there are category theoretic questions about locales. First one notes that `Loc` is complete and cocomplete since the category of frames is known to be complete and cocomplete. This leads to the question how limits and colimits in `Loc` can be computed. As `Loc` is coreflective in `Spec` it follows from general principles that the colimit of a diagram in `Loc` coincides with the colimit computed in `Spec`. Not all colimits in `Spec` are easy to determine. Therefore this answer does not solve all problems concerning colimits. But the situation with limits is more complicated. For example, given two locales X and Y, a product exists in each of the categories `Spec` and `Loc`. But these products are distinct. The product in `Spec` is the same as in `Top`, [2]. (This also follows from the fact that `Spec` is reflective in `Top`.) The product in `Loc` can be constructed in the following way.

First one forms the spectral product $X \times Y$, which is a spectral space, but usually not a locale. Let p_X and p_Y be the projections onto the components. One applies the localic coreflection and obtains the spectral map $R_{X\times Y} : L(X \times Y) \to X \times Y$. Now the domain and the codomains of the maps $p_X \circ R_{X\times Y} : L(X \times Y) \to X$ and $p_Y \circ R_{X\times Y} : L(X \times Y) \to Y$ are locales. But the maps are only spectral and are usually not localic. To complete the construction, there is a largest localic subspace $Z \subseteq L(X \times Y)$ such that $p_X \circ R_{X\times Y}|_Z$ and $p_Y \circ R_{X\times Y}|_Z$ are localic maps. Now the locale Z with these localic maps as projections is the localic product of X and Y.

References

1 M. Dickmann, N. Schwartz, M. Tressl, *Spectral Spaces.* Book manuscript, in preparation
2 M. Hochster, *Prime ideal structure in commutative rings.* Trans. AMS **142**, 43–60 (1969)
3 P. Johnstone, *Stone Spaces.* Cambridge University Press, Cambridge 1986

4 N. Schwartz, *Locales as spectral spaces.* Algebra Universalis **70**, 1–42 (2013)
5 N. Schwartz, *Spectral reflections of topological spaces.* Preprint 2015
6 N. Schwartz, *Localic subspaces and colimits of locales.* Preprint 2015

3.26 Duality for r.e. sets with applications

Vladimir Shavrukov (Utrecht, NL)

We begin with some basic specifics of the dual space $(E^*)_*$ of the lattice E^* of r.e. sets (mod finite). Then we introduce some ultrapower-like models of arithmetic that allow us to have a closer look at the individual points of $(E^*)_*$ (=prime filters of E^*). We show how to apply all of this to obtain a result in the classical recursion-theoretic tradition on r-maximal and hyperhypersimple r.e. sets.

3.27 A Chu-like extension of topological spaces

Paul Taylor (Birmingham, UK)

Main reference P. Taylor, "A Chu-like Extension of Topological Spaces," Slides, 2015.
URL http://paultaylor.eu/slides/15-Dagstuhl.pdf

Dana Scott's *equilogical spaces* provide the best known construction of a cartesian closed category with all finite limits in which the category of sober topological spaces is embedded. An equilogical space $X \equiv A/\sim$ is defined using a partial equivalence relation $\sim$ on the underlying set of an algebraic lattice A. Then the simplest exponential

$$\Sigma^{(A/\sim)} \equiv \Sigma^A/\cong \qquad \text{is given by} \qquad \phi \cong \psi \equiv \forall xy : A.(x \sim y) \Rightarrow (\phi x = \psi y),$$

after which the formulae become more and more complicated. We would like to break up this construction into parts with simpler universal properties in order to understand and compute these exponentials.

The categorical methods that have been used (local cartesian closure, exact completion, embedding in toposes) are in my view *set theory* – tools for the study of *discrete* structures. My objective is to find the *intrinsic* logic of *topology* and higher *computability* theory.

The alternative that we propose here is inspired by Martín Escardó's analogy that equilogical spaces are to topological spaces as complex numbers are to real ones and Steven Vickers' result that the double powerlocale is a monad on the category of locales that reduces to the double exponential $\Sigma^{\Sigma^{(-)}}$ on locally compact spaces. However, we start from the category of *sober topological spaces* because that of locales does not obey the necessary categorical properties.

Recall that *Chu spaces* provide models of Linear Logic with an *involutive negation* that interchanges two components. In our case we replace the identity by a *monad*, for which we obtain a "*square root*" (*cf.* $\sqrt{-1}$), written \$, and a *tensor product* $\otimes$ such that maps

$$X \to \$Y, \qquad Y \to \$X \qquad \text{and} \qquad X \otimes Y \to \Sigma \equiv \$\mathbf{1}$$

are in natural bijection, **1** being the terminal object. In fact $\otimes$ has projections but not in general diagonals.

We may also calculate equalisers, coequalisers and hence (four kinds of) *image factorisation*. That into regular epis and monos defines a *coreflective subcategory*, whose objects do have diagonals, so that $\otimes$ is the categorical product and \$ the exponential. Restricting the subcategory further to those objects that are generated by \$ and finite limits yields general exponentials.

This construction is actually founded on *equideductive logic* rather than point–set topology. With this new tool it may be easier to examine higher exponentials and translate principles from recursion theory to equideductive logic. On the other hand, the Chu-like category may turn out to be a better setting than the cartesian closed one in which to do this study.

In fact, the talk as given diverged somewhat from the slides and explored equideductive logic and its possible relationship to constructive descriptive set theory, since other participants of the seminar such as Matthew de Brecht were interested in this.

3.28 Stone Duality as a Topological Construction

Marcus Tressl (Manchester University, GB)

We present a purely topological construction of Stone Duality passing both ways from a bounded distributive (semi-) lattice to its spectrum and back, using purely topological constructions, such as: Sobrification, Desobrification (passing to the subspace of locally closed points), Hyperspaces and Inverse spaces.

3.29 Monadic second order logic as a model companion of modal logic

Samuel J. van Gool (City University of New York, US)

Joint work of Samuel J. van Gool and Silvio Ghilardi
Main reference S. Ghilardi, S. J. van Gool, "A model-theoretic characterization of monadic second order logic on infinite words," arXiv:1503.08936v1 [math.LO], 2015.
URL http://arxiv.org/abs/1503.08936v1

We exhibit a connection between monadic second order logics and modal logics, making use of the language of first-order model theory; specifically, model companions. Model companions stem from A. Robinson's work on model completions: a model companion of a universal theory T is an extension T^* of T which has the same universal consequences (i.e., T^* is a co-theory of T), but in addition allows for the elimination of quantifier alternations (i.e., T^* is model-complete). Monadic second order logic and modal logic are two different formalisms that can be used to describe the same classes of models, such as infinite words, or infinite trees. Both logics, when interpreted in appropriate power set structures, can actually be viewed as first-order theories. We show in several specific cases that the first-order theory corresponding to MSO is the model companion of the first-order theory corresponding to modal logic. During the talk, I focused on MSO on infinite words, and showed how it can be characterized as the model companion of linear temporal logic with an initial element. I also reported on our ongoing work of extending these results to infinite trees.

3.30 Stone duality without the duality

Steven J. Vickers (University of Birmingham, GB)

The original Stone dualities bring out a wonderful connection between the discrete world of algebra and logic and the continuous world of point-set topology, and they lead to the dual adjunction between frames and topological spaces. Without choice principles the dualities fail, but they still guide the development of point-free topology, replacing point-set structure by algebra. What is left of the original dualities is then less exciting: for example, Stone locales are dual to Boolean algebras. However, they still mark the connection between impredicative world (needing powersets) of frames and the predicative world of frame presentations. It is at this point that duality begins to lose its duality: for how do we define *predicatively* the structures on presentations that correspond to arbitrary maps between the locales? What does it mean to "respect the relations" in the presentations? A general impredicative technique comes from understanding that a classifying topos, or the topos of sheaves over X, is the "geometric mathematics generated by a generic point of X", and that to construct a point of Y in that mathematics is to define a locale map from X to Y. The geometric constraints on the reasoning guarantee that the generic point transformation is continuous. I shall sketch a predicative version of this that is often adequate, in which the geometric structure of Grothendieck toposes is replaced by the arithmetic structure of Joyal's Arithmetic Universes. They have a finitary, coherent internal logic, but also the ability to form types for free algebras including the natural numbers. This allows some infinities (as needed for geometric logic) to be part of the internal structure in a finitary way.

Participants

- Adriana Balan
University Politehnica of Bucharest, RO
- Libor Barto
Charles University – Prague, CZ
- Ulrich Berger
Swansea University, GB
- Nick Bezhanishvili
University of Amsterdam, NL
- Vasco Brattka
Universität der Bundeswehr – München, DE
- Matthew de Brecht
NICT – Osaka, JP
- Max Dickmann
University of Paris VII, FR
- Abbas Edalat
Imperial College London, GB
- Martin H. Escardo
University of Birmingham, GB
- Willem L. Fouché
UNISA – Pretoria, ZA
- Silvio Ghilardi
University of Milan, IT
- Jean Goubault-Larrecq
ENS – Cachan, FR
- Georges Hansoul
University of Liège, BE
- Reinhold Heckmann
AbsInt – Saarbrücken, DE
- Achim Jung
University of Birmingham, GB
- Klaus Keimel
TU Darmstadt, DE
- Dexter Kozen
Cornell University, US
- Andreas Krebs
Universität Tübingen, DE
- Clemens Kupke
University of Strathclyde, GB
- Alexander Kurz
University of Leicester, GB
- Jimmie D. Lawson
Louisiana State University – Baton Rouge, US
- Matteo Mio
ENS – Lyon, FR
- M. Andrew Moshier
Chapman University – Orange, US
- Robert Myers
London, GB
- Dirk Pattinson
Australian National University – Canberra, AU
- Daniela Petrisan
University of Paris VII, FR
- Michael Pinsker
Charles University – Prague, CZ
- Luca Reggio
University of Paris VII, FR
- Giuseppe Rosolini
University of Genova, IT
- Matthias Schröder
TU Darmstadt, DE
- Peter M. Schuster
University of Verona, IT
- Niels Schwartz
Universität Passau, DE
- Victor Selivanov
A. P. Ershov Institute – Novosibirsk, RU
- Vladimir Shavrukov
Utrecht, NL
- Alex Simpson
University of Ljubljana, SI
- Dieter Spreen
Universität Siegen, DE
- Paul Taylor
Birmingham, UK
- Marcus Tressl
Manchester University, GB
- Samuel J. van Gool
City University of New York, US
- Yde Venema
University of Amsterdam, NL
- Steven J. Vickers
University of Birmingham, GB
- Klaus Weihrauch
FernUniversität in Hagen, DE

Report of Dagstuhl Seminar 15442

Approaches and Applications of Inductive Programming

Edited by

José Hernández-Orallo[1], Stephen H. Muggleton[2], Ute Schmid[3], and Benjamin Zorn[4]

1 **Technical University of Valencia, ES,** joralllo@dsic.upv.es
2 **Imperial College London, GB,** s.muggleton@imperial.ac.uk
3 **Universität Bamberg, DE,** ute.schmid@uni-bamberg.de
4 **Microsoft Research – Redmond, US,** zorn@microsoft.com

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 15442 "Approaches and Applications of Inductive Programming". After a short introduction to the state of the art to inductive programming research, an overview of the talks and the outcomes of discussion groups is given.

Seminar October 25–30, 2015 – http://www.dagstuhl.de/15442
1998 ACM Subject Classification I.2.2 Automatic Programming, Program Synthesis
Keywords and phrases inductive program synthesis, end-user programming, probabilistic programming, constraint programming, universal artificial intelligence, cognitive modeling,
Digital Object Identifier 10.4230/DagRep.5.10.89
Edited in cooperation with Fernando Martínez-Plumed

1 Executive Summary

José Hernández-Orallo
Stephen H. Muggleton
Ute Schmid
Benjamin Zorn

Inductive programming research addresses the problem of learning (mostly declarative) programs from incomplete specifications, such as input/output examples, observed traces, or constraints. Beginning in the 1960s, this area of research was initiated in artificial intelligence (AI) exploring the complex intellectual cognitive processes involved in producing program code which satisfies some specification. Furthermore, applications of AI for software engineering are investigated resulting in methodologies and techniques for automating parts of the program development process. Inductive programming can be seen as a very special subdomain of machine learning where the hypothesis space consists of classes of computer programs.

Nowadays, researchers working on inductive programming are distributed over different communities, especially inductive logic programming, evolutionary programming, grammar inference, functional programming, and programming languages and verification. Furthermore, similar approaches are of interest in programming by demonstration applications for end-user programming as well as in cognitive models of inductive learning.

The recent release of FlashFill as a plug-in inductive programming tool for Microsoft Excel is an impressive demonstration that inductive programming research has matured in such a way that commercial applications become feasible. Similarly, the field has attracted widespread interest in computer science as a whole, as illustrated by the recent review article published by the Communications of the ACM [1].

In the seminar, we brought together researchers from different areas of computer science – especially from machine learning, AI, declarative programming, and software engineering – and researchers from cognitive psychology interested in inductive learning as well as in teaching and learning computer programming. Furthermore, participants from industry presented current as well as visionary applications for inductive programming.

We addressed many aspects which partially were identified as relevant topics during the previous Dagstuhl Deminar 13502 (http://www.dagstuhl.de/13502). In particular, we had the following sessions for presentations:

- Session: General techniques, languages and systems for inductive logic and inductive functional programming.
- Session: End-user programming, programming by example and applications
- Session: Program synthesis and transformation
- Cognitive Aspects of Induction

In addition, we had several systems demos and tutorials (some of them 'hands-on'):

- System demo on Metagol.
- Hands-on-Tutorial: The MagicHaskeller Library and Server System
- Tutorial: FlashMeta SDK for creating programming-by-example tools
- Tutorial: Sketch synthesis infrastructure

The seminar also included a DemoFest, where several systems were demonstrated in small groups in a relaxed atmosphere.

The first and second days the following topics were identified and further discussed in working groups during the rest of the seminar:

- Benchmarks, Evaluation, and Applications
- General-Purpose IP Infrastructures and Applicability and Evaluation Criteria for IP Approaches in the Context of AI
- Probabilities in IP

Concluding remarks and future plans

In the final panel discussion the results of the seminar were summarised as well as future plans.

Regarding the seminar, there were several suggestions that topics should be more mixed, instead of grouping them too much into "silo" sessions. About the format of the seminar, there was a general agreement that the change from half a week to one week had been beneficial, and that the DemoFest had been a real success. Indeed, the possibility of having several independent demos earlier in the week and more demo sessions (or DemoFests) was a possible suggestion for subsequent meetings.

The following topics were elaborated about future actions:

- Make the community and the area more visible through tutorials and workshops at major conferences, or summer schools.

- Integrate tools, demos, videos, tutorials and other kinds of material at www.inductive-programming.org/resources.html
- Focus on benchmarks and a common representation language for problems, and use the inductive-programming website to publish the benchmarks. Organise competitions (or hackathons building apps that use the underlying IP engines).
- Revitalise the mailing list (at the moment of writing this report the list is fully operative again, see http://www.inductive-programming.org/)
- Attract people from other areas (e.g., cognitive robotics).
- Change the frequency of the meetings to a 1-year cadence, with perhaps Dagstuhl every other year and a competition or summer school in between.

Overall, the main conclusion can be summarised as the realisation of a very significant progress in techniques and its exploitation in new applications, so it is now time to strengthen the visibility of the IP research and its community, for which this Dagstuhl seminar has served as a lever.

References

1 S. Gulwani, J. Hernández-Orallo, E. Kitzelmann, S. H. Muggleton, U. Schmid, and B. Zorn. *Inductive programming meets the real world.* Communications of the ACM, 58, 11, 2015.

2 Table of Contents

3 Overview of Talks on: Inductive Logic and Inductive Functional Programming

3.1 Meta-Interpretive Learning: achievements and challenges

Stephen H. Muggleton (Imperial College London, GB)

Joint work of Muggleton, Stephen; Cropper, Andrew; Lin, Dianhuan; Tamaddoni-Nezhad, Alireza; Tenenbaum, Joshua; Dechter. Eyal; Ellis, Kevin; Dai, Wang-Zhou
Main reference S. H. Muggleton, D. Lin, A. Tamaddoni-Nezhad, "Meta-interpretive learning of higher-order dyadic datalog: predicate invention revisited," Machine Learning, 100(1):49–73, 2015; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/s10994-014-5471-y
URL http://www.doc.ic.ac.uk/~shm/Papers/metagolD_MLJ.pdf

Meta-Interpretive Learning (MIL) is a recent Inductive Logic Programming technique aimed at supporting learning of recursive definitions. A powerful and novel aspect of MIL is that when learning a predicate definition it automatically introduces sub-definitions, allowing decomposition into a hierarchy of reuseable parts. MIL is based on an adapted version of a Prolog meta-interpreter. Normally such a meta-interpreter derives a proof by repeatedly fetching first-order Prolog clauses whose heads unify with a given goal. By contrast, a meta-interpretive learner additionally fetches higher-order meta-rules whose heads unify with the goal, and saves the resulting meta-substitutions to form a program. This talk will overview theoretical and implementational advances in this new area including the ability to learn Turing computabale functions within a constrained subset of logic programs, the use of probabilistic representations within Bayesian meta-interpretive and techniques for minimising the number of meta-rules employed. The talk will also summarise applications of MIL including the learning of regular and context-free grammars, learning from visual representions with repeated patterns, learning string transformations for spreadsheet applications, learning and optimising recursive robot strategies and learning tactics for proving correctness of programs. The talk will conclude by pointing to the many challenges which remain to be addressed within this new area.

References

1 A. Cropper and S. H. Muggleton. Learning efficient logical robot strategies involving composable objects. In *Proceedings of the 24th International Joint Conference Artificial Intelligence (IJCAI 2015)*, pp. 3423–3429, 2015.
2 W.-Z. Dai, S. H. Muggleton, and Z.-H. Zhou. Logical vision: Meta-interpretive learning for simple geometrical concepts. In Short Paper Proceedings of the 25th International Conference on Inductive Logic Programming. National Institute of Informatics, Tokyo, 2015.
3 S. H. Muggleton, D. Lin, and A. Tamaddoni-Nezhad. Meta-interpretive learning of higher-order dyadic datalog: Predicate invention revisited. *Machine Learning*, 100(1):49–73, 2015.
4 D. Lin, E. Dechter, K. Ellis, J. B. Tenenbaum, and S. H. Muggleton. Bias reformulation for one-shot function induction. In *ECAI*, 2014.

3.2 Towards Probabilistic Logic Program Synthesis

Luc De Raedt (KU Leuven, BE)

Probabilistic logic programs combine the power of a programming language with a possible world semantics; they are typically based on Sato's distribution semantics [5] and they have been studied for over twenty years now. They have recently been extended towards defining continuous distributions and dynamics, which enables their use in robotics and perception [1]. The talk shall briefly introduce these formalisms and then present some progress on synthesising such probabilistic programs from examples, both in the discrete and the continuous case. For the discrete case, I shall report on our results in applying ProbFOIL [5] to the problem of machine reading in CMU's Never Ending Language Learning system. ProbFOIL is an extension of the traditional rule-learning system FOIL for use with the distribution semantics. For the continuous case, I shall present our ongoing work in learning affordances in robotics, where the goal is to learn the conditions under which actions can be applied on particular objects [2, 3].

References

1 B. Gutmann, I. Thon, A. Kimmig, M. Bruynooghe, and L. De Raedt. The magic of logical inference in probabilistic programming. Theory and Practice of Logic Programming, 2011(11), pp. 663–680.

2 D. Nitti, T. De Laet, L. De Raedt, A particle filter for hybrid relational domains. in Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2013, pp. 2764–2771.

3 B. Moldovan, P. Moreno, M. van Otterlo, J. Santos-Victor, L. De Raedt, Learning relational affordance models for robots in multi-object manipulation tasks. In Proc. IEEE International Conference on Robotics and Automation, ICRA 2012, pp. 4373–4378.

4 L. De Raedt, A. Dries, I. Thon, G. Van den Broeck, M. Verbeke, Inducing Probabilistic Relational Rules from Probabilistic Examples. In Proc. International Joint Conference on AI, IJCAI 2015, in press.

5 T. Sato, A Statistical Learning Method for Logic Programs with Distribution Semantics. In Proc. 12th International Conference on Logic Programming, ICLP 1995, pp. 715–729

3.3 Learning Complex Sequential Patterns with Progol

Michael Siebers (Universität Bamberg, DE)

Joint work of Schmid, Ute; Siebers, Michael

The analysis of facial expressions is a new application domain for ILP. Facial expressions may be described by symbols, called action units, which represent movements in the face. Then, a facial expression is a sequence of action units possibly occurring in parallel. I will present first results on learning a generalized representation of facial expressions of pain using Progol. Additionally, I will provide ideas for additional background knowledge to improve results.

3.4 RuleML as a Declarative Language for Inputs, Outputs, and Background Knowledge in Inductive Programming

Harold Boley (University of New Brunswick at Fredericton, CA)

RuleML is a system of families of languages for Web rules connecting related efforts such as: SWRL, SWSL, and RIF (W3C); SBVR, PRR, API4KP, and OntoIOp (OMG); Common Logic (ISO); and LegalRuleML (OASIS).

Inductive Programming can use RuleML to represent inputs, outputs, and background knowledge, since:

- it combines relational/logical and equational/functional representations, e.g. Prolog-like relations and functions defined as oriented (conditional) equations;
- its canonical XML format allows modular (Relax NG-schema) validation of (challenge/benchmark/...-)library entries, each w.r.t. the most precise language;
- RuleML/XML also allows (XSLT) transformation, e.g. to favorite presentation syntaxes of library users and to other XML-based standards;
- its engines permit the execution/querying of induced programs/rulebases;
- its canonical syntax, validation methods, interoperation techniques, as well as, e.g., operational and model-theoretic semantics can support the evaluation of Inductive Programming systems.

4 Overview of Talks on: End-user Programming, Programming by Example and Applications

4.1 Applications to Data Wrangling

Sumit Gulwani (Microsoft Corporation – Redmond, US)

Main reference S. Gulwani, "Programming by Examples (and its applications in Data Wrangling)," In Verification and Synthesis of Correct and Secure Systems 2016, Marktoberdorf Lecture Notes, IOS Press, to appear; pre-print available from author's webpage.
URL http://research.microsoft.com/en-us/um/people/sumitg/pubs/pbe16.pdf

99% of computer end users do not know programming and struggle with repetitive tasks. Inductive synthesis can revolutionize this landscape by enabling end users to automate repetitive tasks using examples. In order to realize this potential, we need to apply inductive synthesis to the right set of application domains. Data wrangling turns out to be a killer application area for inductive synthesis.

Data is the new oil. Evolution of digital revolution, social media, cloud computing, IoT has led to production of massive amounts of digital data. This data is the new currency of the digital world since it can help drive business decisions, advertising, recommendation systems, etc. Data wrangling refers to the tedious process of transforming data from its raw format to a more structured form that is amenable for drawing insights. It is estimated that data scientists spend 80% of their time wrangling with data. Inductive synthesis can enable easier and faster data wrangling. We have developed inductive synthesis tools for assisting with various data wrangling activities including string/number/date transformations (FlashFill), extraction of structured data from semi-structured log files or webpages (FlashExtract), and formatting or table layout transformations (FlashRelate). FlashFill has been released as an Excel 2013 feature, while FlashExtract has been released as the ConvertFrom-string Powershell cmdlet and the custom field extraction capability in Azure OMS.

Practical deployment of inductive synthesis tools require addressing an important challenge associated with inductive synthesis systems, namely resolving ambiguity in the example based specification. We address this challenge using two key ideas: (i) machine learning based ranking techniques to predict an intended program from within the set of programs that are consistent with the examples provided by the user, (ii) user interaction models (including program navigation and active-learning based conversational clarification) that communicate actionable information to the user to help resolve ambiguity in the example based specification.

4.2 Generic vs. domain specific approaches to IP – How good does Igor perform on FlashFill problems?

Ute Schmid (Universität Bamberg, DE)

Joint work of Kitzelmann, E.; Hofmann, M.; Hofmann, J., Schmid, U.
Main reference U. Schmid, E. Kitzelmann, "Inductive rule learning on the knowledge level," Cognitive Systems Research, 12(3–4):237–248, 2011.
URL http://dx.doi.org/10.1016/j.cogsys.2010.12.002

Approaches to inductive programming (IP) learn (declarative) (recursive) programs from a small number of (positive) input/output examples. These approaches are generic, that is, if the given problem can be solved by operations produced by a program, the program can be induced (within some restricted time and space and if the program can be expressed by the underlying language restriction given by an implicit or explicit program scheme). On the other hand, applications of IP in the domain of enduser programming are typically based on a domain specific language which allows to extract suitable examples from observing interactions of users with the given software. If IP is generic, it should be able to be applied 'from the shelf' to different domains and consequently also to the domains investigated in enduser programming.

In my talk I demonstrated how the inductive functional programming system Igor performs on programming problems, planning problems, number series problems, XSLT peroblems, and string transformation problems. The last domain is where the Excell plug-in Flashfill performs in a very impressive way. Looking at the problem specifications given to Igor, we can see that for some domains these are quite simple and natural, for other domains (such as string transformations in Excell) they are quite clumsy. Futhermore, to present the examples in a form suitable for inductive generalization, the selection of examples as well as their representation is crucial. So I claim that extracting suitable representations from raw data such as observations of user interactions or images is the bottleneck of IP.

References

1 Flener, P., & Schmid, U. (2010). Inductive programming. In C. Sammut & G. Webb (Eds.), *Encyclopedia of machine learning*, pp. 537–544. Springer.
2 Hofmann, J., Kitzelmann, E., & Schmid, U. (2014). Applying inductive program synthesis to induction of number series a case study with igor2. *KI 2014: Advances in Artificial Intelligence*, pp. 25–36. Springer.

3 Hofmann, M., Kitzelmann, E., & Schmid, U. (2009). A unifying framework for analysis and evaluation of inductive programming systems. *Proceedings of the Second Conference on Artificial General Intelligence (AGI-09, Arlington, Virginia, March 6-9 2009)*, pp. 55–60. Amsterdam: Atlantis Press.
4 Kitzelmann, E., & Schmid, U. (2006). Inductive synthesis of functional programs: An explanation based generalization approach. *Journal of Machine Learning Research*, *7*, pp. 429–454.
5 Schmid, U., & Kitzelmann, E. (2011). Inductive rule learning on the knowledge level. *Cognitive Systems Research*, *12*, pp. 237–248.

4.3 Resolving Ambiguity in Programming By Example

Rishabh Singh (Microsoft Research – Redmond, US)

Since Programming By Examples is inherently ambiguous, there are typically a large number of programs that conform to the given set of examples especially if the hypothesis space is large. In this talk, we present two way to tackle this ambiguity problem. The first approach uses machine learning techniques to learn a ranking function form training data to efficiently rank the set of learnt programs and select the top ranked program. We have implemented this technique in the FlashFill PBE system, which reduces the average number of examples needed to learn the desired program from 4.17 to 1.48 on 175 benchmarks. The second approach adds probabilistic semantics to the underlying DSL for learning robust transformations on semantic data type strings to handle ambiguity.

4.4 Cooperative Programming: Integrating Software Synthesis with the Live Paradigm

Ruzica Piskac (Yale University, US)

Live programming is an emerging paradigm that is promising a vast change in the techniques used to develop modern software. A live programming environment allows a programmer to immediately see the effects of changes to a program on its outputs and effectively eliminates the edit-run-debug cycle that dominates programming workflows today.

We propose an approach that seeks to develop new synthesis techniques that make use of the real-time feedback loop provided by a live programming environment. We propose a system that will allow a user to track a set of examples and synthesize subroutines to fit that set. When combined with fault localization techniques, programmers will be able to quickly find incorrect sections of code and initiate repairs that leverage information gleaned from the provided examples.

4.5 Deductive Techniques for Synthesis from Inductive Specifications

Sumit Gulwani (Microsoft Corporation – Redmond, US)

Main reference S. Gulwani, "Programming by Examples (and its applications in Data Wrangling)," In Verification and Synthesis of Correct and Secure Systems 2016, Marktoberdorf Lecture Notes, IOS Press, to appear; pre-print available from author's webpage.
URL http://research.microsoft.com/en-us/um/people/sumitg/pubs/pbe16.pdf

We propose two key ideas for designing efficient algorithms for inductive synthesis problems. (i) restrict the search space to an appropriate domain-specific language that provides the right set of abstractions that enable readability and balanced expressivity (i.e., expressive enough to capture a wide range of tasks, but restricted enough to allow efficient search). (ii) design a domain-specific search algorithm using the divide-and-conquer paradigm that reduces the problem of synthesizing an expression of a given kind that satisfies a given specification into sub-problems that refer to sub-expressions or sub-specifications.

The problem reduction logics can be refactored into domain-independent parts and operator-specific parts. We leverage this observation to develop a general framework that allows construction of efficient inductive synthesizers from a mere description of the domain-specific language and inverse properties of operators that are used in the underlying DSL.

4.6 Supervision by observation using inductive programming

José Hernández-Orallo (Technical University of Valencia, ES)

Joint work of Monserrat, Carlos; Hernández-Orallo, José

Despite the fact that many professional or personal tasks we do every day are still very repetitive, the technologies to automate them are still incipient. A less idealistic, but still challenging short and mid-term goal is a scenario where the tasks are still performed by humans but supervised by machines. In other words, we aim at automatically *supervising* these tasks, an area of potentially high impact.

For instance, the training of new skills and the human errors during their execution entail a high cost that could be reduced by automated supervision systems indicating when, where and how a protocol error takes place, in a comprehensible way for the user or the professional. These systems should assimilate the procedures from expert knowledge and the observation of a few correct examples for the procedure without the need of being coded by a programmer.

I present the embryo project "SuPERVaSION", which aims at exploring the pervasive use of inductive programming to address this problem domain. The potentiality of the project is planned to be demonstrated by their application to automatic skill supervision in minimally invasive surgery.

Acknowledgements. This work has been partially supported by the EU (FEDER) and the Spanish MINECO under grant TIN 2013-45732-C4-1-P and Generalitat Valenciana PROMETEOII/2015/013.

References

1 S. Gulwani, J. Hernández-Orallo, E. Kitzelmann, S. H. Muggleton, U. Schmid, and B. Zorn. *Inductive programming meets the real world.* Communications of the ACM, 58, 11, 2015.

2 J. Hernández-Orallo. *Deep knowledge: Inductive programming as an answer.* Technical Report, Seminar 13502, Dagstuhl, 2013.

3 C. Monserrat, A. Lucas, J. Hernández-Orallo, and M. José Rupérez. *Automatic supervision of gestures to guide novice surgeons during training.* Surgical endoscopy, 28(4):1360–1370, 2014.

4 C. Monserrat, M. J. Rupérez, M. Alcañiz, J. Mataix. *Markerless monocular tracking system for guided external surgery.* Computerized Medical Imaging and Graphics, Elsevier, pp. 1–8, 2014.

4.7 Software is not fragile

William B. Langdon (University College London, GB)

Joint work of Mark Harman, Justyna Petke, Estibaliz Aldecoa-Otalor, Phil Cunningham, Matthew J. Arno, Westley Weimer

Main reference W. B. Langdon, J. Petke, "Software is Not Fragile," in Proc. of the 2015 World e-Conference Complex Systems Digital Campus (CS-CD'15), to appear; pre-print available from the author's webpage.

URL http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/papers/langdon_2015_csdc.pdf

URL http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/gp-code/grammar_test_ilp.tar.gz

The talk, as requested, concentrated upon insights gained, in my case from working on genetic programming. These are summarised on slide 3 of my talk: work on industrial strength languages, focus search, evolve patches (change to C program source), evolve source code v. machine code, ensure many patches/mutants compile, software resilient to mutation, choose receptive domain, separate fitness from validation, evolution exploits fitness, present results on a slide, e.g. source code, ...

I then concentrated upon recent work which counters the "folk wisdom" with results from experiments from a number of real programs which show software is robust to many (albeit not all) random changes. I also included recent results that show more than a 70x speed up on a real program when customised to the task in hand with the use of genetic improvement using mutations of the C++ source code.

During the seminar I was also asked about the in silco spread of bacterial genes in the Human reference genome, genetic programming, human competetive GI and learning context free language benchmarks (benchmarks are available as tar file from http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/gp-code/grammar_test_ilp.tar.gz)

Trying all simple changes (first order mutations) to executed C, C++ and CUDA source code shows software engineering artefacts are more robust than is often assumed. Of those that compile, up to 89percent run without error. Indeed a few mutants are improvements. Program fitness landscapes are smoother. Analysis of these programs, a parallel nVidia GPGPU kernel, all CUDA samples and the GNU C library shows many lines of code and integer values are repeated and may follow Zipf's law.

We show genetic improvement of programs (GIP) can scale by evolving increased performance in a widely-used and highly complex 50 000 line system. GISMOE found code that is 70 times faster (on average) and yet is at least as good functionally. Indeed it even gives a small semantic gain.

References

1 William B. Langdon and Justyna Petke. Software is not fragile. In Paul Bourgine and Pierre Collet, editors, *Complex Systems Digital Campus E-conference, CS-DC'15*, Proceedings in Complexity, page Paper ID: 356. Springer, September 30 – October 1 2015. Invited talk, Forthcoming.

2 William B. Langdon and Mark Harman. Optimising existing software with genetic programming. *IEEE Transactions on Evolutionary Computation*, 19(1):118–135, February 2015.

3 William B. Langdon. Mycoplasma contamination in the 1000 genomes project. *BioData Mining*, 7(3), 29 April 2014. Highly accessed.

4 Estibaliz Aldecoa-Otalora, William B. Langdon, Phil Cunningham, and Matthew J. Arno. Unexpected presence of mycoplasma probes on human microarrays. *BioTechniques*, 47(6):1013–1016, December 2009.

5 Justyna Petke, Mark Harman, William B. Langdon, and Westley Weimer. Using genetic improvement and code transplants to specialise a C++ program to a problem class. In Miguel Nicolau, Krzysztof Krawiec, Malcolm I. Heywood, Mauro Castelli, Pablo Garcia-Sanchez, Juan J. Merelo, Victor M. Rivas Santos, and Kevin Sim, editors, *17th European Conference on Genetic Programming*, volume 8599 of *LNCS*, pp. 137–149, Granada, Spain, 23-25 April 2014. Springer.

6 Riccardo Poli, William B. Langdon, and Nicholas Freitag McPhee. *A field guide to genetic programming*. Published via `http://lulu.com` and freely available at http://www.gp-field-guide.org.uk, 2008. (With contributions by J. R. Koza).

7 William B. Langdon. Genetically improved software. In Amir H. Gandomi, Amir H. Alavi, and Conor Ryan, editors, *Handbook of Genetic Programming Applications*, chapter 8. Springer. Forthcoming.

5 Overview of Talks on: Program synthesis and transformation

5.1 Synthesis of service specifications: Applying evolutionary algorithms to On-the-fly computing

Lorijn van Rooijen (Universität Paderborn, DE)

The aim of our research project on On-the-fly (OTF) Computing is to develop concepts and techniques for automatic on-the-fly composition of individualized IT services out of base services that are available on world-wide markets. One part of this project deals with the problem of obtaining, in a user-friendly way, requirements specifications for the services that users would want to acquire. To facilitate the user, our objective is to automatically synthesize such requirements specifications from sample specifications that are provided by the user. These sample specifications may be incomplete and imprecise and might even contain contradictions. We plan to employ evolutionary algorithms to synthesize comprehensive requirements specifications that do justice to the sample specifications by extending them in a natural way.

5.2 Inductive Program Synthesis for Bidirectional Transformations

Janis Voigtländer (Universität Bonn, DE)

Joint work of Tobias Gödderz, Helmut Grohne, Janis Voigtländer
Main reference J. Voigtländer, "Ideas for connecting inductive program synthesis and bidirectionalization," in Proc. of ACM SIGPLAN 2012 Workshop on Partial Evaluation and Program Manipulation (PEPM'12), pp. 39–42, ACM, 2012.
URL http://dx.doi.org/10.1145/2103746.2103757

Bidirectional transformations are a mechanism for preserving the consistency of (at least) two related data structures. A classical incarnation is the view-update problem, which has received considerable attention from programming language research. Specifically, bidirectionalization techniques attempt to automatically produce a well-behaving backwards transformation from a given forwards transformation. Well-behavedness is expressed by certain roundtripping laws. We explore the use of inductive program synthesis for bidirectionalization. The main idea is to use the roundtripping laws, and possibly additional constraints derived from supposed programmer intentions, as a source of input/output examples for the backwards transformation. We report on recent experiments using the Igor-II system.

5.3 D3: Data-Driven Disjunctive Abstraction

Hila Peleg (Technion – Haifa, IL)

Joint work of Hila Peleg, Sharon Shoham, Eran Yahav]

We address the problem of computing an abstraction for a set of examples, which is precise enough to separate them from a set of counterexamples. The challenge is to find an over-approximation of the positive examples that does not represent any negative example. Conjunctive abstractions (e.g., convex numerical domains) and limited disjunctive abstractions, are often insufficient, as even the best such abstraction might include negative examples. One way to improve precision is to consider a general disjunctive abstraction. We present D3, a new algorithm for learning general disjunctive abstractions. Our algorithm is inspired by widely used machine-learning algorithms for obtaining a classifier from positive and negative examples. In contrast to these algorithms which cannot generalize from disjunctions, D3 obtains a disjunctive abstraction that minimizes the number of disjunctions. The result generalizes the positive examples as much as possible without representing any of the negative examples. We demonstrate the value of our algorithm by applying it to the problem of data-driven differential analysis, computing the abstract semantic difference between two programs. Our evaluation shows that D3 can be used to effectively learn precise differences between programs even when the difference requires a disjunctive representation.

5.4 Program transformation using examples

Gustavo Soares (Universidade Federal – Campina Grande, BR)

Joint work of Reudismam Rolim, Rohit Gheyi, Sumit Gulwani

During software evolution, developers may have to apply similar but not identical changes to different locations in the program. Manually performing these changes is time-consuming and error-prone. We then propose an inductive programming technique for automating repetitive changes using examples. First, the developer provides examples of locations to be changed. The technique learns a program that finds similar code fragments throughout the developers' code. The developer then applies the transformation to one or more locations. Based on the examples of the desired transformation, the technique learns a program that applies the transformation and transforms all locations identified in step 1. We identified 68 scenarios where developers performed systematic changes and our approach was able to automate these tasks.

6 Overview of Talks on: Cognitive Aspects of Induction

6.1 Learning new Domains Without the Help of Engineers

Claes Strannegård (Chalmers University of Technology – Göteborg, SE)

It has been postulated that the goal of artificial general intelligence is to create general intelligence at the human level and beyond [4]. To get anywhere near that goal one needs to construct versatile agents that can adapt to a wide range of environments without any human intervention. In natural nervous systems, reinforcement learning is a powerful mechanism that enables organisms to adapt to different environments and survive there [3]. In artificial systems, reinforcement learning has been used as the basis of relatively versatile agents, e.g. for robotic locomotion across different anatomies and for gaming across different arcade games [1].

In this talk I will discuss how more versatile systems might be constructed, e.g. systems that can both learn to move like a snake and do simple mathematics. I will sketch a system whose main parts are (i) a long-term memory in the form of a Markov Decision Process based on transparent networks that develops dynamically [2], (ii) a learner consisting of rules for evolving the long-term memory by adding and removing memory structures, and (iii) a decision-maker for planning and acting.

References

1 Volodymyr Mnih and others. *Human-level control through deep reinforcement learning*. Nature, 518(7540):529–533, 2015
2 Claes Strannegård, Simone Cirillo, and Johan Wessberg. *Emotional Concept Formation*. Proc. of the 8th Conf. on Artificial General Intelligence. pp. 166–176, Springer, 2015
3 Yael Niv. *Reinforcement learning in the brain*. Journal of Mathematical Psychology. 53(3):139–154, 2009
4 Ben Goertzel and Cassio Pennachin. *Artificial General Intelligence*. Vol. 2, Springer, 2007.

6.2 Knowledge acquisition with forgetting: an incremental and developmental view

Fernando Martínez-Plumed (Technical University of Valencia, ES)

Joint work of Fernando Martínez-Plumed, Cèsar Ferri, José Hernández-Orallo, María José Ramírez-Quintana:
Main reference F. Martínez-Plumed, C. Ferri, J. Hernández-Orallo, M. J. Ramírez-Quintana, "Knowledge acquisition with forgetting: an incremental and developmental setting," Adaptive Behaviour, 23(5):283–299, 2015.
URL http://dx.doi.org/10.1177/1059712315608675

The development of a knowledge discovery system that is meant to be incremental and cumulative is not an easy task: the use of background knowledge and the consolidation of new knowledge is one of the conspicuous problems in the understanding and creation of cognitive systems, and the management of more lifelong knowledge discovery systems. Knowledge acquisition, as the process of abstracting knowledge from facts and other knowledge, cannot be understood as a naive accumulation of what is being learnt. Identifying the balance between remembering and forgetting is the key to abstraction in the human brain (creation of memories and knowledge) [1] and, therefore, this could be also applied when developing knowledge acquisition AI systems.

In this talk I present the work in [4], an incremental, lifelong view of knowledge acquisition which tries to improve task after task by determining what to keep, what to consolidate and what to forget, overcoming *The Stability-Plasticity* dilemma [2]. This framework is designed to combine any rule-based inductive engine (which learns new rules) with a deductive engine (which derives a coverage graph for all rules) and integrates them into a lifelong learner through the use of a hierarchical knowledge assessment structure (*coverage graphs*) and by introducing several MML-based [3] metrics. The metrics are not only used to forget some of the worst rules, but also to set a consolidation process to promote those selected rules to the knowledge base, which is also mirrored by a demotion system.

Acknowledgements. This work has been partially supported by the EU (FEDER) and the Spanish MINECO under grant TIN 2013-45732-C4-1-P and Generalitat Valenciana PROMETEOII/2015/013.

References

1 Quiroga, R. Q. *Concept cells: the building blocks of declarative memory functions*. Nature Reviews Neuroscience, vol. 13, pp. 587–597, 2012.
2 Carpenter, G. A. and Grossberg, S. *The art of adaptive pattern recognition by a selforganizing neural network*. Computer, vol. 21, no. 3, pp. 77–88, 1988.
3 Wallace, C. S. and Boulton, D. M. *An information measure for classification*.The Computer Journal, vol. 11, no. 2, pp. 185–194, 1968.
4 F. Martínez-Plumed and C. Ferri and J. Hernández-Orallo and M. J. Ramírez-Quintana *Knowledge acquisition with forgetting: an incremental and developmental setting*. Adaptive Behavior, 23(5):283–299, 2015.

6.3 A Visual Language for Solving Bongard Problems

Frank Jäkel (Universität Osnabrück, DE)

Joint work of Depeweg, S.; Rothkopf, C.; Jäkel, F.

More than 50 years ago [1] introduced a set of 100 vision problems, now known as Bongard problems, as a test-bed for visual pattern recognition. Although these problems are well known in the cognitive science and AI communities only moderate progress has been made towards solving a substantial subset of them. The approach we present here extracts standard visual features as a basic visual vocabulary. We introduce a formal language that allows representing complex visual concepts and relations using this vocabulary. Finally, using Bayesian inference on the space of concepts formulated in this visual language, we compare the concepts with high posterior probability to the solutions formulated by Bongard himself when designing his problems. We find good agreement for a sizeable fraction of the problems.

References

1 M. Bongard. *Pattern Recognition.* Spartan Books, New York, 1970.

6.4 The Artificial Jack of All Trades: The Importance of Generality in Approaches to AI

Tarek R. Besold (Free University of Bozen-Bolzano, IT)

Joint work of Besold, Tarek R.; Schmid, Ute
Main reference T. R. Besold, U. Schmid, "The Artificial Jack of All Trades: The Importance of Generality in Approaches to Human-Level Artificial Intelligence", in Proc. of the 3rd Annual Conf. on Advances in Cognitive Systems (ACS'15), Article No. 18, 18 pages, Cognitive Systems Foundation, 2015.
URL http://www.cogsys.org/papers/ACS2015/article18.pdf

We advocate the position that research efforts working towards solving human-level AI necessarily have to rely on general mechanisms and (models of) cognitive capacities, with domain-specific systems or task-dependent approaches only being of minor help towards the final goal. We revisit psychological research on intelligence and the application of psychometric methods in AI, before discussing IGOR2 (implementing program learning) and HDTP (implementing generalisation-based analogy-making) under the light of the previous considerations as examples of systems respectively implementing a general mechanism or modeling a general capacity. In closing, we summarise our considerations, point out three characteristics we consider suitable candidates for serving as generally recommendable properties of HLAI systems, and motivate why AI as a field could greatly profit from closer interaction with Inductive Programming (and vice versa).

7 System Demonstrations

7.1 Hands-on Tutorial for Hacking MagicHaskeller

Susumu Katayama (University of Miyazaki, JP)

This presentation was a hands-on tutorial of MagicHaskeller, that is a powerful inductive functional programming system for Haskell focusing on practical use. Although the presentation material was prepared for the users who are interested in using its API, the actual presentation focused on its web interface and was mainly for those who were not familiar with it.

8 Working groups

8.1 Benchmarks, Evaluation, and Applications

Umair Zafrulla Ahmed (Indian Institute of Technology – Kanpur, IN), Harold Boley (University of New Brunswick at Fredericton, CA), José Hernández-Orallo (Technical University of Valencia, ES), Petra Hofstedt (TU Cottbus, DE), Frank Jäkel (Universität Osnabrück, DE), William B. Langdon (University College London, GB), Fernando Martinez-Plumed (Technical University of Valencia, ES), Martin Möhrmann (Universität Osnabrück, DE), Hila Peleg (Technion – Haifa, IL), Maria José Ramirez Quintana (Technical University of Valencia, ES), Gustavo Soares (Universidade Federal – Campina Grande, BR), Armando Solar-Lezama (MIT – Cambridge, US), Lorijn van Rooijen (Universität Paderborn, DE), Janis Voigtländer (Universität Bonn, DE), and Benjamin Zorn (Microsoft Research – Redmond, US)

The goal of this working group is to increase the visibility and impact of research in the areas of inductive programming as applied to important practical problems. We want to facilitate a broad effort to coordinate and increase communication among researchers, identify important research directions, and show the utility of IP in practice across important application domains. To achieve this goal, we identify two important sub-goals: creating a repository of problem description and links to open-source systems (which has been already initiated at http://inductive-programming.org/) and to host competitions focused on specific tasks and domains to highlight the significant progress being made on practical applications of IP.

Application areas: In our discussions we identified that there are many emerging areas where practical applications of IP are both significant and commercially valuable. These include general tests of intelligence, data manipulation, education, programming (code transformations, test generation, etc.), grammar discovery, and visual reasoning. Our hope is that we can add problem sets in all of these areas in the coming year that will significantly broaden the set of problems that are already present in our repository.

Common representations: One of the challenges of defining a set of shared problem instances is the challenge of defining the problems in such a way that that they can be used across systems. Ideally, we would have a way to encode the problem (capturing the input/output pairs, etc.). We also have to encode characteristics of the resulting artifact (e.g., it might be a program in a domain specific language). Finally, encoded in the solution and also the search used in finding the solution is the embedded background knowledge that allows the solution to closely match that a person might construct in solving the problem. It is especially difficult to tease out this background knowledge in some systems because it can be encoded in a domain-specific language or even as a set of constraints that exist in the solver system that is used.

Competitions: One of the key areas we wanted to increase activity in during the coming years is competitions that highlight important application areas for IP and allow the systems that are emerging to demonstrate how effective they are. Armando shared some of his experiences running the SyGuS synthesis competition for 2 consecutive years. He pointed out a number of lessons that include: avoid narrowing the community, allow multiple winners and different categories, have in mind for the first competition a list of participants that will definitely enter, think carefully about the problems ahead of time.

We considered specific opportunities for connecting a competition to a particular meeting and KDD2016 was discussed but the deadline is too soon. We decided that focusing the first competition on data wrangling would limit the scope to an important area and allow a number of different approaches to be used (string manipulation is common to a number of the existing systems). The plan would be to change the topic of the competition each year to diversify the applications explored. ICDM 2016 in Barcelona is a possible alternative venue we are considering. http://www.cs.uvm.edu/~icdm/

Evaluation: Another important dimension of benchmarks and evaluation is what metrics are applied to determine if the solution is "good". In many cases, the "naturalness" of a solution to a human is an important requirement. We explored increasing the use of crowdsourcing as a way to evaluate aspects of a solution that are more difficult to measure using objective techniques. One starting point for this is to create a catalog of the existing metrics that different researchers are using and understand if those metrics are domain specific or general.

Community: One of the important challenges in building collections of benchmarks, maintaining repositories and keeping them fresh, and organizing competitions is that none of these activities directly leads to new results or publications. As a result, a community is needed that gets benefit from shared infrastructure and sees synergy in the related efforts. One possible goal in this area might be to publish a paper that captures a set of benchmarks in a way that makes it much easier for subsequent research to build on that result. For example, the Dicapo benchmarks for Java are highly used and cited in the Java research community. One of the things we plan to do going forward is maintaining regular interactions among the attendees of the seminar interested in this topic via regular Skype meetings. We also discussed other forms of social media and how to share and communicate important results in this research area.

8.2 General-Purpose IP Infrastructures and Applicability and Evaluation Criteria for IP Approaches in the Context of AI

Ute Schmid (Universität Bamberg, DE), Tarek R. Besold (Free University of Bozen-Bolzano, IT), Andrew Cropper (Imperial College London, GB), Sumit Gulwani (Microsoft Corporation – Redmond, US), Petra Hofstedt (TU Cottbus, DE), Susumu Katayama (University of Miyazaki, JP), William B. Langdon (University College London, GB), and Stephen H. Muggleton (Imperial College London, GB)

The topic of inductive programming (IP) was discussed in the context of the general goals and quality criteria of "classical" artificial intelligence (AI) research (i.e., research aiming at developing human-level AI systems). The initial proposition was that general purpose approaches to IP – such as inductive logic programming (ILP) and inductive functional programming (IFP) – are more compatible with the goals of AI research than specialized, domain-specific approaches. However, the general scientific goal of AI research to formalize human-level intelligence by means of general algorithms is in practice complemented by the application oriented engineering goal to provide efficient and reliable algorithmic solutions for problems of interest for industry and end-users. Nevertheless, taking into account insights about human cognition is also of relevance for application-oriented AI in order to provide for adequate and comprehensible human-computer interaction and human-understandable and -interpretable processing approaches.

Generality of mechanisms (for reasoning, induction, problem solving, etc.) is hypothesized as hallmark of human cognition. That is, it is assumed that the same processing principles guide, for example, solving geometric induction problems such as Raven progressive matrices, induction of number series, induction of replacement patterns for string transformations, or induction of a general routine to solve Tower of Hanoi or sorting of lists. While ILP and IFP are such general mechanism, the domain specific approach of the Excel plug-in Flashfill for string transformations demonstrates the enormous power of specialized approaches within their restricted domains. An engineering approach to reach both the performance of specialized systems as well as the coverage of a broader set of domains could be to provide a generator mechanism which allows to construct domain-specific solutions.

Within AI research machine learning is the area of research most closely related to IP: In IP as well as in standard ML, the main goal is to provide algorithms for inductive generalization over incomplete information. In standard ML, training examples are typically presented in the form of feature vectors, learning algorithms are robust to noise, and in order to ensure suitable generalization and avoid overfitting it is not desirable that the learned hypothesis covers all training examples. In contrast, in IP, training examples are presented in the form of input/output relations and in addition background knowledge can be provided. Furthermore, since the learned hypothesis is a (declarative) program, it is necessary that all input/output examples are treated correctly. While most of standard ML approaches are blackbox learners, IP approaches are whitebox learners. That is, the generated hypothesis is a symbolic and structured representation which is comprehensible by humans. The big difference between a standard ML approach such as deep learning neural networks and IP is, that IP addresses deep comprehension! In consequence, we propose that comprehensibility might be a suitable performance criterium for the output of ML/IP approaches.

A promising line of research in IP might be to make systems which generate output which is (even more) compatible with human-to-human communication. That is, the output of an IP system should have high comprehensibility. Comprehensibility implies natural interaction, safety and trust. As a first step in this direction, we need to characterize what defines the comprehensibility of system output (such as names, concepts, action sequences, programs, texts, theories, designs, pictures, emotional responses and so on). Consequently, a new IP system could be designed to use comprehensibility as inductive bias. To determine the comprehensibility of a learned program, the characteristics of comprehensibility could be learned in form of a mapping from characteristics of the candidate hypotheses to the degree or probability of understanding. Understanding could be empirically measured as probability that the subject can provide the correct classification (or action) for arbitrary input based on the provided program. A first step could be attempting to machine learn definitions of comprehensibility from labelled examples provided by humans. This could be done in the form of setting comprehension tests, similar in spirit to those set for texts in schools, in which the human subjects were asked to classify the implied consequences of various machine generated hypotheses. The expected accuracy of such tests would then be taken as a proxy measure for the degree of comprehensibility of individual hypotheses with respect to the given group of humans tested.

One characteristic of making a program more comprehensible than another one could be the use of predicate (or function) invention as used by ILP systems such as Golem, Progol or Metagol and by the IFP system Igor. Invented predicates/functions introduce a new level of abstraction and allow to structure a program by using higher-level concepts in the program body which are defined separately. A simple example is a Prolog definition of the concept grandfather/2 with and without predicate invention:

```
; without predicate invention
p(X,Y) ← father(X,Z), father(Z,Y)
p(X,Y) ← father(X,Z), mother(Z,Y)
p(X,Y) ← mother(X,Z), mother(Z,Y)
p(X,Y) ← mother(X,Z), father(Z,Y)

;with predicate invention
p(X,Y) ← p1(X,Z), p1(Z,Y)
p1(X,Y)← father(X,Y)
p1(X,Y)← mother(X,Y)
```

As outcome of the discussion group, it was planned to design and conduct an empirical study about comprehensibility of Prolog programs with versus without invented predicates.

8.3 Probabilities in Inductive Programming

Rishabh Singh (Microsoft Research – Redmond, US), Luc De Raedt (KU Leuven, BE), Cesar Ferri Ramirez (Technical University of Valencia, ES), Frank Jäkel (Universität Osnabrück, DE), Maria José Ramirez Quintana (Technical University of Valencia, ES), Michael Siebers (Universität Bamberg, DE), Armando Solar-Lezama (MIT – Cambridge, US), and Christina Zeller (Universität Bamberg, DE)

Our discussion group focused on the role of probabilistic analysis in Inductive programming. We identified three main ways in which probabilities can be used: 1) Modelling priors of the programs in the hypothesis space to specify a program P1 is more likely than another program P2, 2) Using probabilistic programs for inductive reasoning so that a program can generate a distribution of outputs for modeling noise, and finally 3) a unified setting where both noise and prior can be modeled together using a joint distribution. We then identified some applications in inductive programming that can be enabled using such probabilistic analysis. One of the key new applications was synthesizing programs in settings where there is some inherent ambiguity in inputs and outputs, e.g. human-generated data, weather/GPS and sensor data. The probabilistic analysis can also increase the robustness of previous deterministic inductive programming approaches by using better priors for the hypothesis space and handling noisy data, e.g. a robust FlashFill learning algorithm. Finally, another interesting application was in formalizing the heuristics used in synthesis approaches.

We then contrasted probabilistic analysis in Inductive Programming with Machine Learning and Probabilistic Programming. As compared to Machine Learning, the probabilistic IP approaches can handle both hard and soft constraints, can learn human-understandable programs instead of complex high-dimensional models, and provide more control over the hypothesis space. Probabilistic Programming, on the other hand, can be used to formulate our problem, but they are not yet scalable enough to handle real-world IP problems. There is also a strong bias towards sampling based methods in Probabilistic Programming and lesser emphasis on combinatorial approaches that are more common in IP. Finally, we discussed some ways in which to perform a good evaluation of Probabilistic IP systems. The common measures of number of examples needed to learn a program and the scalability of the learning algorithm are similar to the traditional IP systems, but are harder to evaluate in our setting. Another interesting measure to evaluate in probabilistic IP systems is to verify how faithfully the learning algorithm is conforming to the probability distributions. Finally, we discussed some potential ways to evaluate the prior models for the hypothesis space and input-output distributions.

Participants

- Umair Zafrulla Ahmed
Indian Institute of Technology – Kanpur, IN
- Tarek R. Besold
Free Univ. of Bozen-Bolzano, IT
- Harold Boley
University of New Brunswick at Fredericton, CA
- Andrew Cropper
Imperial College London, GB
- Luc De Raedt
KU Leuven, BE
- Cesar Ferri Ramirez
Technical Univ. of Valencia, ES
- Sumit Gulwani
Microsoft Corporation – Redmond, US
- José Hernández-Orallo
Technical Univ. of Valencia, ES
- Petra Hofstedt
TU Cottbus, DE
- Frank Jäkel
Universität Osnabrück, DE
- Susumu Katayama
University of Miyazaki, JP
- William B. Langdon
University College London, GB
- Fernando Martinez-Plumed
Technical Univ. of Valencia, ES
- Martin Möhrmann
Universität Osnabrück, DE
- Stephen H. Muggleton
Imperial College London, GB
- Hila Peleg
Technion – Haifa, IL
- Ruzica Piskac
Yale University, US
- Maria José Ramirez Quintana
Technical Univ. of Valencia, ES
- Ute Schmid
Universität Bamberg, DE
- Michael Siebers
Universität Bamberg, DE
- Rishabh Singh
Microsoft Res. – Redmond, US
- Gustavo Soares
Universidade Federal – Campina Grande, BR
- Armando Solar-Lezama
MIT – Cambridge, US
- Claes Strannegård
Chalmers University of Technology – Göteborg, SE
- Lorijn van Rooijen
Universität Paderborn, DE
- Janis Voigtländer
Universität Bonn, DE
- Christina Zeller
Universität Bamberg, DE
- Benjamin Zorn
Microsoft Res. – Redmond, US