Report from Dagstuhl Seminar 17192

Human-Like Neural-Symbolic Computing

Edited by

Tarek R. Besold[1], Artur d'Avila Garcez[2], and Luis C. Lamb[3]

1 **Universität Bremen, DE,** tbesold@uni-bremen.de
2 **City, University of London, GB,** aag@soi.city.ac.uk
3 **Federal University of Rio Grande do Sul, BR,** luislamb@acm.org

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 17192 "Human-Like Neural-Symbolic Computing", held from May 7th to 12th, 2017. The underlying idea of Human-Like Computing is to incorporate into Computer Science aspects of how humans learn, reason and compute. Whilst recognising the relevant scientific trends in big data and deep learning, capable of achieving state-of-the-art performance in speech recognition and computer vision tasks, limited progress has been made towards understanding the principles underlying language and vision understanding. Under the assumption that neural-symbolic computing – the study of logic and connectionism as well statistical approaches – can offer new insight into this problem, the seminar brought together computer scientists, but also specialists on artificial intelligence, cognitive science, machine learning, knowledge representation and reasoning, computer vision, neural computation, and natural language processing. The seminar consisted of contributed and invited talks, breakout and joint group discussion sessions, and a hackathon. It was built upon previous seminars and workshops on the integration of computational learning and symbolic reasoning, such as the Neural-Symbolic Learning and Reasoning (NeSy) workshop series, and the previous Dagstuhl Seminar 14381: Neural-Symbolic Learning and Reasoning.

Seminar May 7–12, 2017 – http://www.dagstuhl.de/17192
1998 ACM Subject Classification F.1.1 Models of Computation: Self-modifying machines, I.2 Artificial Intelligence, I.2.10 Vision and Scene Understanding, I.2.4 Knowledge Representation Formalisms and Methods, I.2.6 Learning, I.2.6 Learning: Connectionism and neural nets, I.2.7 Natural Language Processing
Keywords and phrases Deep Learning, Human-like computing, Multimodal learning, Natural language processing, Neural-symbolic integration
Digital Object Identifier 10.4230/DagRep.7.5.56

1 Executive Summary

Luis C. Lamb
Tarek R. Besold
Artur d'Avila Garcez

The underlying idea of Human-Like Computing is to incorporate into Computer Science aspects of how humans learn, reason and compute. Recognising the relevance of the scientific trends in big data, data science methods and techniques have achieved industrial relevance in a number of areas, from retail to health, by obtaining insight from large data collections. Notably, neural networks have been successful and efficient at large-scale language modelling, speech recognition, image, video and sensor data analysis [3, 12, 15]. Human beings, on the other hand, are excellent at learning from very few data examples, capable of articulating explanations and resolving inconsistencies through reasoning and communication [7, 9, 16, 17].

Despite the recent impact of deep learning, limited progress has been made towards understanding the principles and mechanisms underlying language and vision understanding. Under this motivation, the seminar brought together not only computer scientists, but also specialists in artificial intelligence (AI), cognitive science, machine learning, knowledge representation and reasoning, computer vision, neural computation and natural language processing. In particular, the methodology of neural-symbolic computation [4, 7, 12], which can offer a principled interface between the relevant fields, especially symbolic AI and neural computation, was adopted in an attempt to offer a new perspective of reconciling large-scale modelling with human-level understanding, thus building a roadmap for principles and applications of Human-Like Neural-Symbolic Computing.

The techniques and methods of neural-symbolic computation have already been applied effectively to a number of areas, leading to developments in deep learning, data science and human-like computing [3]. For instance, neural-symbolic integration methods have been applied to temporal knowledge evolution in dynamic scenarios [6, 10, 14], action learning in video understanding [10], uncertainty learning and reasoning [1], argument learning in multiagent scenarios [7, 8], hardware and software verification and learning [2], ontology learning [13] and distributed temporal deep learning in general, with several applications in computer science [2, 10, 14].

Specifically, in this Dagstuhl Seminar we aimed at: (i) building better bridges between symbolic and sub-symbolic reasoning and learning, and between big data and human-like learning; (ii) comparative analyses and evaluations of the explanatory capacity of language modelling tools and techniques; (iii) design and applications of knowledge extraction methods and techniques towards life-long learning and transfer learning.

The seminar consisted of contributed and invited talks, breakout and joint group discussion sessions, and scientific hackathons. After each presentation or discussion session, open problems were identified and questions were raised. The area is clearly growing in importance, given recent advances in Artificial Intelligence and Machine Learning. In particular, the need for explainability in AI clearly poses relevant questions to learning methodologies, including deep learning. In summary, the main research directions identified by participants are:

- Explainable AI: The recent success of deep learning in vision and language processing, associated with the growing complexity of big data applications has led to the need for *explainable AI models*. In neural-symbolic computing, rule extraction, interpretability, comprehensibility leading to the development of integrated systems, are one of the principled alternatives to lead these efforts [5, 7], as discussed in the Explainability hackathon. Furthermore, the concept of modularity in multimodal learning in deep networks is crucial to the development of the field and can help achieve knowledge extraction (as identified in [5, 6]) which can result in the development of effective knowledge extraction methods towards explainable AI, as discussed in the deep learning with symbols hackathon.
- Hybrid Cognitive Architectures: The development of Cognitive Architectures capable of simulating and explaining aspects of human cognition also remains an important research endeavour. Some cognitive architectures typically consider symbolic representations, whereas others employ neural simulations. The integration of these models remains a challenge and there are benefits on integrating the accomplishments of both paradigms, as identified in the cognitive architectures hackathon.

- Statistical Relational Learning: Logic Tensor Networks (LTNs) [11] provides a model that integrates symbolic knowledge (encoded as first-order logic relations) and subsymbolic knowledge (represented as feature vectors). The LTNs enable the representation of relational knowledge infusion into deep networks, and knowledge completion and distilling through querying the networks. There remains a number of challenges in integrating, explaining and computing symbolic knowledge in deep networks. Both LTNs [11] and Connectionist Modal and Temporal Logics [6, 7, 14] offer effective alternatives towards these research challenges, as explored in the LTN hackathon.

The seminar builds upon previous seminars and workshops on the integration of computational learning and symbolic reasoning, such as the Neural-Symbolic Learning and Reasoning (NeSy) workshop series, and the previous Dagstuhl Seminar 14381: Neural-Symbolic Learning and Reasoning [5].

References

1 Tarek R. Besold, Artur S. d'Avila Garcez, Keith Stenning, Leendert W. N. van der Torre, Michiel van Lambalgen: Reasoning in Non-probabilistic Uncertainty: Logic Programming and Neural-Symbolic Computing as Examples. *Minds and Machines*, 27(1): 37-77, 2017.

2 Rafael V. Borges, Artur S. d'Avila Garcez, Luis C. Lamb. *Learning and Representing Temporal Knowledge in Recurrent Networks*. IEEE Trans. Neural Networks and Learning Systems 22(12):2409-2421, Dec. 2011. IEEE Press.

3 A. d'Avila Garcez, Tarek R. Besold, Luc De Raedt, Peter Földiak, Pascal Hitzler, Thomas Icard, Kai-Uwe Kühnberger, Luis C. Lamb, Risto Miikkulainen, Daniel L. Silver. Neural-Symbolic Learning and Reasoning: Contributions and Challenges. Proceedings of the AAAI Spring Symposium on Knowledge Representation and Reasoning: Integrating Symbolic and Neural Approaches. Stanford Univ., March 2015, pp. 18-21, AAAI press, 2015.

4 Artur S. d'Avila Garcez, Dov M. Gabbay, Krysia Broda. *Neural-Symbolic Learning System: Foundations and Applications*. Springer-Verlag, New York, Inc., USA, 2002.

5 A. S. d'Avila Garcez, Marco Gori, Pascal Hitzler, Luís C. Lamb: Neural-Symbolic Learning and Reasoning (Dagstuhl Seminar 14381). Dagstuhl Reports 4(9): 50-84, 2015.

6 Artur S. d'Avila Garcez and Luis C. Lamb. A Connectionist Computational Model for Epistemic and Temporal Reasoning. *Neural Computation*, 18(7):1711-1738, 2006.

7 Artur S. d'Avila Garcez, Luis C. Lamb and Dov M. Gabbay. *Neural-Symbolic Cognitive Reasoning*. Springer Publishing Company, 2009. ISBN: 9783540732457.

8 A. d'Avila Garcez, D.M. Gabbay and Luis C. Lamb: A neural cognitive model of argumentation with application to legal inference and decision making. Journal of Applied Logic, 12(2):109-127, 2014.

9 M. de Kamps and F. van de Velde. From neural dynamics to true combinatorial structures. Behavioral and Brain Sciences, 20, pp. 88-108, 2006.

10 H.L. de Penning, Artur S. d'Avila Garcez, Luis C. Lamb, John-Jules Ch. Meyer: *A Neural-Symbolic Cognitive Agent for Online Learning and Reasoning*. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI 2011. Pages 1653-1658, 2011.

11 Ivan Donadello, Luciano Serafini, and Artur d'Avila Garcez. *Logic Tensor Networks for Semantic Image Interpretation*. Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017.

12 Barbara Hammer, Pascal Hitzler (Eds.): Perspectives of Neural-Symbolic Integration. Studies in Computational Intelligence 77, Springer 2007.

13 P. Hitzler, S. Bader, and A. d'Avila Garcez. *Ontology learning as a use case for neural-symbolic integration*. In Proc. Workshop on Neural-Symbolic Learning and Reasoning, NeSy'05 at IJCAI-05.

14 Luis C. Lamb, R.V. Borges and A.S. d'Avila Garcez. *A connectionist cognitive model for temporal synchronisation and learning*. Proc. AAAI Conference on Artificial Intelligence AAAI-07, pages 827-832, AAAI Press, 2007.

15 Y. LeCun, Y. Bengio, G. Hinton. Deep Learning. *Nature*, 521 (7553): 436-444, 2015.

16 G.F. Marcus, S. Vijayan, S.B. Rao, and P.M. Vishton. Rule learning by seven-month-old infants. *Science*, 283(5398):77–80, 1999.

17 K. Stenning and M. van Lambalgen. *Human reasoning and Cognitive Science*. MIT Press, Cambridge, MA, 2008.

18 Alan M. Turing. Computing Machinery and Intelligence, *Mind*, LIX (236): 433–460, 1950.

2 Table of Contents

3 Overview of Talks

3.1 Conceptual Spaces: A Bridge Between Neural and Symbolic Representations?

Lucas Bechberger (Universität Osnabrück, DE)

The cognitive framework of conceptual spaces [1] attempts to bridge the gap between symbolic and subsymbolic AI by proposing an intermediate conceptual layer based on geometric representations. A conceptual space is a high-dimensional space spanned by a number of quality dimensions representing interpretable features. Convex regions in this space correspond to concepts. Abstract symbols can be grounded by linking them to concepts in a conceptual space whose dimensions are based on subsymbolic representations.

The framework of conceptual spaces has been highly influential in the last 15 years within cognitive science and cognitive linguistics. It has also sparked considerable research in various subfields of artificial intelligence, ranging from robotics and computer vision over the semantic web and ontology integration to plausible reasoning.

Although this framework provides means for connecting concepts from the symbolic layer to numeric information from the subsymbolic layer, it does not yet provide an automated way to do so: In practical applications, both the mapping from the subsymbolic to the conceptual layer (i.e., how to map observations to points) and the mapping from the conceptual layer to the symbolic layer (i.e., how to map regions to symbols) need to be handcrafted by a human expert.

After introducing the conceptual spaces framework in more detail, I argue that we can use machine learning in order to learn these mappings: I propose to use representation learning (namely the InfoGAN framework [2]) for learning the dimensions of a conceptual space from unlabeled data. Moreover, I suggest to use an incremental clustering algorithm to discover meaningful regions in a conceptual space that can then give rise to abstract symbols.

References

1 Peter Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. MIT Press, 2000
2 Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel, *InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets*. Advances in Neural Information Processing Systems, 2016.

3.2 Comprehensible ILP

Tarek R. Besold (Universität Bremen, DE)

Joint work of Ute Schmid, Christina Zeller, Stephen Muggleton, Alireza Tamaddoni-Nezhad, Tarek R. Besold
Main reference Schmid U., Zeller C., Besold T., Tamaddoni-Nezhad A., Muggleton S. (2017) How Does Predicate Invention Affect Human Comprehensibility?. In: Cussens J., Russo A. (eds) Inductive Logic Programming. ILP 2016. Lecture Notes in Computer Science, vol 10326. Springer, Cham
URL http://dx.doi.org/10.1007/978-3-319-63342-8_5

During the 1980s Michie defined Machine Learning in terms of two orthogonal axes of performance: predictive accuracy and comprehensibility of generated hypotheses. Since predictive accuracy was readily measurable and comprehensibility not so, later definitions in the 1990s, such as that of Mitchell, tended to use a one-dimensional approach to Machine Learning based solely on predictive accuracy, ultimately favoring statistical over symbolic Machine Learning approaches. In this talk I want to discuss a definition of comprehensibility of hypotheses which can be estimated using human participant trials.

To illustrate the proposed approach, we will have a look at two recent experiments testing human comprehensibility of ILP-learned logic programs. The first experiment tested human comprehensibility with and without predicate invention, with results indicating that comprehensibility is affected not only by the complexity of the presented program but also by the existence of anonymous predicate symbols. The second experiment showed that a state-of-the-art ILP system can support humans effectively beyond their own capabilities in a relational concept learning task.

3.3 Using evidence accumulation to bridge the gap between neural networks and symbolic cognitive control

Jelmer Borst (University of Groningen, NL) and Terrence Stewart

I will discuss how we used large-scale spiking neural networks to simulate associative recognition. Associative recognition is the important ability to learn that two items co-occur. Although detailed symbolic models exist that account for behavior, fMRI, and EEG data, it remains unclear how associative recognition is performed at the neural level. To investigate this, we used the Neural Engineering Framework to simulate associative recognition with spiking neural networks that can process symbols and coordinate cognition through the basal ganglia [2]. Because the resulting neural network model is very complex (> 500,000 neurons) we use magnetoencephalographic (MEG) data to constrain the model [1]. The model matches data in occipital, temporal, prefrontal, and motor cortices, and shows how the associative recognition process could be implemented in the human brain.

References

1 Jelmer P. Borst, Avniel S. Ghuman, & John R. Anderson *Tracking cognitive processing stages with MEG: A spatio-temporal model of associative recognition in the brain*. NeuroImage, 141:416–430, 2016.
2 Chris Eliasmith, Terrence C. Stewart, Xuan Choo, Trevor Bekolay, Travis DeWolf, Yichuan Tang, & Daniel Rasmussen *A Large-Scale Model of the Functioning Brain*. Science, 338:1202–1205, 2012.

3.4 Neural-Symbolic Systems for Human-Like Computing

Artur d'Avila Garcez (City, University of London, GB)

An increasing number of researchers now accept that the recent success of deep learning needs to be combined with an understanding of its reach, scope and limitations so that the area can continue to move forward towards Artificial General Intelligence (AGI). Neural-symbolic computing has been making foundational contributions in this direction for a long time now [1, 3], as it is concerned with the interplay between learning and reasoning, and the combination of rich symbolic AI formalisms with efficient neural computation frameworks. The applications of AGI are vast as it permits humans to improve their own performance at various tasks from interacting with (explainable) learning systems. AGI also permits systems to learn from other systems as part of what became known as transfer learning. Finally, such general learning systems will need to become verifiable against certain properties such as safety and trust as they become commonplace in our daily lives [2].

In "Neural-Symbolic Systems for Human-Like Computing", I presented an overview of the neural-symbolic methodology, including knowledge insertion (aka infusion), learning, reasoning and knowledge extraction (aka distilling) from neural networks. This methodology enables the integration of symbolic and sub-symbolic AI in a principled way. I then exemplified the neural-symbolic cycle by applying it to software model verification and adaptation [5], and to the use of neural networks for the run-time monitoring of software system properties specified in linear temporal logic [4]. I concluded by discussing ways in which neural-symbolic systems may contribute to the research on human-like computing, notably in the areas of representation change, bridging the gap between high and low-level learning and reasoning, memory and forgetting, comprehensibility and explanation, and learning from only a few examples.

References

1. Garcez, Artur S. d'Avila, Gabbay, Dov M. and Broda, Krysia B., *Neural-Symbolic Learning System: Foundations and Applications*. Springer-Verlag, New York, Inc., USA, 2002. ISBN: 1852335122.
2. Garcez, Artur S. d'Avila and Lamb, Luis C., *Reasoning About Time and Knowledge in Neural-symbolic Learning Systems*. In Proc. 16th International Conference on Neural Information Processing Systems NIPS'03, MIT Press, Cambridge, MA, USA, 2003.
3. Garcez, Artur S. d'Avila, Lamb, Luis C. and Gabbay, Dov M., *Neural-Symbolic Cognitive Reasoning*. Springer Publishing Company, 2008. ISBN: 9783540732457.
4. Alan Perotti, Artur S. d'Avila Garcez and Guido Boella, *Neural-symbolic monitoring and adaptation*. In Proc. International Joint Conference on Neural Networks, IJCNN'15, Killarney, Ireland, 2015.
5. Borges, Rafael V., d'Avila Garcez, Artur, Lamb, Luis C., *Learning and Representing Temporal Knowledge in Recurrent Networks*. IEEE Trans. Neural Networks and Learning Systems 22(12):2409–2421, Dec. 2011. IEEE Press.

3.5 Dynamics for the Neural Blackboard Architecture

Marc de Kamps (University of Leeds, GB)

Humans surpass computers when it comes to understanding. Computers excel in object recognition or sentence parsing, but when it comes to extracting meaning, humans do a lot better. We partly understand why this is so: neural representations are grounded [4]: neural activity derives meaning from where it occurs relative to the sensory pathways or other brain areas. This is a major advantage of computer representations consisting of strings of binary numbers, which have no intrinsic meaning. The network knowledge representation of the brain makes it relatively straightforward to understand how associations occur and how attractor networks can implement memory. This implementation raises fundamental questions about how compositional representations are represented, however. It is not easy to understand how the same concept can occur twice in a different context within the same scene or sentence [6]. It is not easy to understand how abstract relationships are realized by the brain and how they can be applied to concrete instances, when an agent must resolve its constituents in order to extract meaning To address this problem, we proposed the Neural Blackboard Architecture (NBA), a biologically plausible cognitive architecture that aims to explain these capabilities in terms of neural dynamics [5]. We have reached the point where we can make predictions about sequences of neural activity in language processing, and validate them against experimental results. Preliminary results of this work have been reported in abstract form [2].

For such a program, realistic models of neural dynamics at the population level are essential. I will report recent progress in population density techniques (e.g. [1, 3]), developed as part of the Human Brain Project, to show how large groups of spiking neurons can be modeled efficiently at the population level, without using thousands of model neuron instances. I will also show preliminary modeling results of this technique applied to the NBA. When four word sentences are compared with a baseline consisting of four words not constituting a sentence, there is a marked difference in the neural signal. Future work will make the comparison to experimental results. A big advantage of the technique presented here is that neural models can be exchanged easily without affecting the structure of the network, which makes it possible to vary biological assumptions in the model (e.g. the presence of adaptation). The population density technique implementation is available as Open Source software on http://miind.sf.net

Acknowledgment: The population modeling work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 720270 (Human Brain Project).

References

1 Y.M. Lai and M. de Kamps. Population density equations for stochastic processes with memory kernels, Phys. Rev. E 95, 062125, 2017.
2 Martin Perez-Guevara, Marc De Kamps, Christophe Pallier. Blackboard Architecture simulation with simple biological networks captures the behavior of diverse neuroimaging measurements during language processing, in NIPS Workshop Representation Learning in Artificial and Biological Neural Networks, 2016.
3 de Kamps M, A Generic Approach to Solving Jump Diffusion Equations with Applications to Neural Populations, https://arxiv.org/abs/1309.1654, 2013.
4 F. van der Velde, M. de Kamps. Compositional connectionist structures based on in situ grounded representations. Connect Sci, 23 (2), pp. 97-107, 2011.
5 M. de Kamps, F. van de Velde. From neural dynamics to true combinatorial structures. Behavioral and Brain Sciences, 20 , pp. 88-108, 2006.
6 R. Jackendoff. Foundations of Language: Brain, Meaning, Grammar, Evolution, Oxford University Press, 2003.

3.6 On Explainability in Machine Learning

Derek Doran (Wright State University – Dayton, US)

Joint work of Ning Xie, Md Kamruzzaman Sarker, Eric Nicols, Pascal Hitzler, Michael Raymer, Derek Doran

Enabling explainability in statistical machine learning is a fast emerging topic. The development is rooted in the idea that statistical ML models are widely perceived (perhaps unfairly) as "black box", where input/output mappings are performed by mathematical operations that are not easily interpretable or explainable. But (many) statisticians and ML researchers disagree: one can argue that regression models are interpretable, and any model with linear decision boundaries for classification imply a set of decision making "rules". This presentation introduces a hierarchy of explainability that statistical ML algorithms may be classified into: interpretable, explainable, and reasonable, and includes recent examples of each. It then introduces a new approach for enabling explanations, and potentially higher level reasoning, for neural networks. The approach involves: 1) localized representations of concepts through regularization; 2) applying an inference engine that explains these localized activations by discriminating between inputs activating the localization and those that don't.
Acknowledgements: This work is supported by the Ohio Federal Research Network.

3.7 Tackling Commonsense Reasoning Benchmarks

Ulrich Furbach (Universität Koblenz-Landau, DE)

Joint work of Ulrich Furbach, Claudia Schon
Main reference U. Furbach, C. Schon, "Commonsense Reasoning Meets Theorem Proving", in Proc. of the 14th German Conference on Multiagent System Technologies (MATES 2016), LNCS, Vol. 9872, pp. 3–17, Springer, 2016.
URL https://doi.org/10.1007/978-3-319-45889-2_1

The area of commonsense reasoning aims at the creation of systems able to simulate the human way of rational thinking. This talk describes the use of automated reasoning methods for tackling commonsense reasoning benchmarks. For this we use a benchmark suite introduced in literature. Our goal is to use general purpose background knowledge without domain specific hand coding of axioms, such that the approach and the result can be used as well for other domains in mathematics and science. Furthermore, we discuss the modeling of normative statements in commonsense reasoning and in robot ethics.

3.8 What do Neural Networks need in order to generalize?

Raquel Garrido Alhama (University of Amsterdam, NL) and Willem Zuidema

In an influential paper, reporting on a combination of artificial language learning experiments with babies, computational simulations and philosophical arguments, [1] claimed that connectionist models cannot account for human success at learning tasks that involved generalization of abstract knowledge such as grammatical rules. This claim triggered a heated debate, centered mostly around variants of the Simple Recurrent Network model [2]. In this paper, we revisit this unresolved debate and analyze the underlying issues from a different perspective. We argue that, in order to simulate human-like learning of grammatical rules, a neural network model should not be used as a tabula rasa, but rather, the initial wiring of the neural connections and the experience acquired prior to the actual task should be incorporated into the model. We present two methods that aim to provide such initial state: a manipulation of the initial connections of the network in a cognitively plausible manner (concretely, by implementing a "delay-line" memory), and a pre-training algorithm that incrementally challenges the network with novel stimuli. We implement such techniques in an Echo State Network [3], and we show that only when combining both techniques the ESN is able to succeed at the grammar discrimination task suggested by Marcus et al.

References

1 G. F. Marcus, S. Vijayan, S.B. Rao, and P.M. Vishton. Rule learning by seven-month-old infants. *Science*, 283(5398):77–80, 1999.
2 Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.
3 Herbert Jaeger. *The "echo state" approach to analysing and training recurrent neural networks*. Technical report, German National Research Center for Information Technology, 2001.

3.9 Understanding Neural Networks through Background Knowledge

Pascal Hitzler (Wright State University – Dayton, US), Derek Doran (Wright State University – Dayton, US), Maryam Labaf, Md Kamruzzaman Sarker, Michael Raymer, and Ning Xie

The ever increasing prevalence of publicly available structured data on the World Wide Web enables new applications in a variety of domains. In this presentation, we provide a conceptual approach that leverages such data in order to explain the input-output behavior of trained artificial neural networks. We apply existing Semantic Web technologies in order to provide an experimental proof of concept. The presentation starts by investigating the case of propositional rule extraction as a base case, carrying past results over to incorporate background knowledge. In the second part, we incorporate knowledge graphs and ontologies and show how the DL-Learner symbolic machine learning system can be used to generate explanations which take such background knowledge into account.

Acknowledgements. This work was supported by the Ohio Federal Research Network project *Human-Centered Big Data*.

References

1 Maryam Labaf and Pascal Hitzler, Propositional Rule Extraction from Neural Networks under Background Knowledge. In: Proceedings of the Twelfth International Workshop on Neural-Symbolic Learning and Reasoning, London, UK, July 2017. To appear.
2 Md Kamruzzaman Sarker and Ning Xie and Derek Doran and Michael Raymer and Pascal Hitzler, Explaining Trained Neural Networks with Semantic Web Technologies: First Steps. In: Proceedings of the Twelfth International Workshop on Neural-Symbolic Learning and Reasoning, London, UK, July 2017. To appear.

3.10 Processing Hierarchical Structure with RNNs

Dieuwke Hupkes (University of Amsterdam, NL)

Joint work of Willem Zuidema, Sara Veldhoen, Dieuwke Hupkes

We investigate how recurrent neural networks can learn and process languages h hierarchical, compositional semantics. To this end, we define the simple but nontrivial artificial task of processing nested arithmetic expressions and study whether recurrent neural networks, which process these expressions incrementally, can learn to compute their meaning. We show that a neural network architecture with gating (the Gaed Recurrent Unit) performs well on this task: the network learns to predict the outcome of the arithmetic expressions with high accuracy, although performance deteriorates with increasing length. To analyse what strategy the recurrent network applies, visualisation techniques are not sufficient. Therefore, we develop an approach where we formulate and test hypothess on what strategies such networks might be following. For each hypothesis, we derive predictions about featurs of the hidden state representations at each time step, and train 'diagnostic' classifiers to test those predictions. Our results indicate that the networks follow a strategy similar to our hypothesised 'cumultative strategy'.

3.11 How do people build and comprehend ontologies?

Caroline Jay (University of Manchester, GB)

Joint work of Caroline Jay, Nicolas Matentzoglu, Markel Vigo, Robert Stevens
Main reference N. Matentzoglu, M. Vigo, C. Jay, R. Stevens, "Making Entailment Set Changes Explicit Improves the Understanding of Consequences of Ontology Authoring Actions", in Proc. of the 20th Int'l Conf. on Knowledge Engineering and Knowledge Management (EKAW 2016), LNCS, Vol. 10024, pp. 432–446, Springer, 2016.
URL https://doi.org/10.1007/978-3-319-49004-5_28

Ontologies have become an important feature of many fields, including science, commerce and medicine. They support the representation of a domain's knowledge for use in a range of applications. Ontologies describe classes of objects, and the relationships between them. When in logical form, these descriptions can be used to draw conclusions about the knowledge represented; this information can be used both for developing the ontology itself, and in an application. An ontology can become very large and highly complex, however, and a single change can have many unanticipated effects (both desirable and undesirable) that are difficult for a human engineer to comprehend [2]. Accuracy is important: if the representation is incorrect, the value of any knowledge resulting from reasoning over it is undermined.

At the University of Manchester, via the EPSRC-funded project 'WhatIf: Answering "What if. . . " questions for Ontology Authoring' (EP/J014176/1), we have been exploring how people build ontologies, and comprehend the information within them. This work shows that people use the class hierarchy (within the Protege tool) as an external memory, spending almost half their time looking or interacting with it when working on an ontology. They also use the reasoner and the inferred class hierarchy to check the consequences of their modelling actions [3].

Understanding how a particular action has affected an ontology is extremely difficult, as there may be multiple, unanticipated changes that have implications for a complex series of axioms. To ameliorate the situation, we designed and tested the Inference Inspector, a tool that shows the most relevant consequences of modelling actions to the ontology developer. Making entailment set changes explicit significantly improves the speed and accuracy of ontology development, as well as improving author comprehension of the effects of his or her changes [1].

References

1 Nicolas Matentzoglu, Markel Vigo, Caroline Jay and Robert Stevens. *Making Entailment Set Changes Explicit Improves the Understanding of Consequences of Ontology Authoring Actions*. Knowledge Engineering and Knowledge Management – 20th International Conference EKAW 2016, Bologna, Italy, November 19–23, 2016.

2 Markel Vigo, Sam Bail, Caroline Jay and Robert Stevens. *Overcoming the pitfalls of ontology authoring: Strategies and implications for tool design.* International Journal of Human-Computer Studies, 2014.

3 Markel Vigo, Caroline Jay and Robert Stevens. *Constructing Conceptual Knowledge Artefacts: Activity Patterns in the Ontology Authoring Process.* CHI, 2015.

3.12 Human-like software engineering: bridging the human-machine translation gap

Caroline Jay (University of Manchester, GB)

Joint work of Caroline Jay, Robert Haines, Robert Stevens
Main reference C. Jay, R. Haines, M. Vigo, N. Matentzoglu, R. Stevens, J. Boyle, A. Davies, C. Del Vescovo, N. Gruel, A. Le Blanc, D. Mawdsley, D. Mellor, E. Mikroyannidi, R. Rollins, A. Rowley, J. Vega Hernandez, "Identifying the challenges of code/theory translation: report from the Code/Theory 2017 workshop", Research Ideas and Outcomes 3: e13236, Pensoft Publishers, 2017.
URL http://dx.doi.org/10.3897/rio.3.e13236

Engineering software is a challenging endeavour. Development processes are incrementally improving, allowing us to construct increasingly complex artefacts, yet software continues to contain errors, or behave in unforeseen ways. This is partly to due to the 'unknown unknowns' introduced by a changing external environment, but it is also because algorithms often fail to work as expected: the formal representations underlying machine computation are frequently at odds with the heuristics used by the human brain. Whilst this is an issue across software development, it is particularly apparent in scientific software engineering, where the purpose of the code implementation is to represent, precisely, a scientific entity, process or system, and a given programming language is often an inadequate means of doing this [1]. The result is a *human-machine translation gap*.

Whilst empirical software engineering is a thriving field, using a theory of human cognition based on empirical observation to drive system development is not a standard approach; progress has occurred primarily through craft-based iteration, rather than rigorous empirical study. Nevertheless, observation of the programming process has demonstrated its potential to result in huge technological advances. A notable example of this is locality of reference, a principle uncovered when trying to ascertain how to page data in and out of memory, which has gone on to touch virtually every aspect of modern systems. The *theory of locality* describes how data relevant to the current context of a running program is grouped together locally in space and time. Using this model to page data in and out of memory dramatically improved system performance, and laid the foundations for a multitude of other optimisations, including efficient caching, network configuration and search engine ranking. There are two key aspects of this early work that demonstrate the value of a 'human-like' approach to software engineering: the first is that locality is a product of the way people write programs, rather than being due to any underlying constraints of the computing system; the second is that it was discovered empirically – paging algorithms were ineffective for many years before the principle of locality was discovered through systematic observation of the programming process [2].

Advances in hardware, such as parallel processing, have yet to achieve their full potential, as we struggle to translate serial human-written programs onto distributed architectures. Automated programming offer a means to reduce and repair errors, but even with machine-written programs, human input to a system means a bottleneck will always remain. As we move into the era of quantum computing and beyond, a true understanding of how our minds map themselves onto the machines we create is a vital component of achieving a step change in the creation and performance of software.

References

1 Caroline Jay and Robert Haines and Markel Vigo and Nicolas Matentzoglu and Robert Stevens and Jonathan Boyle and Alan Davies and Chiara Del Vescovo and Nicolas Gruel and Anja Le Blanc and David Mawdsley and Dale Mellor and Eleni Mikroyannidi and Richard Rollins and Andrew Rowley and Julio Vega Hernandez. *Identifying the challenges of code/theory translation: report from the Code/Theory 2017 workshop*. DOI: 10.3897/rio.3.e13236. Research Ideas and Outcomes, Pensoft Publishers, 2017.

2 Peter Denning. *The locality principle*. Communications of the ACM, 2007.

3.13 (Human-Like) Anecdotes with IBM's Watson/Bluemix Services

Kai-Uwe Kühnberger (Universität Osnabrück, DE)

Joint work of Kai-Uwe Kühnberger, Gordon Pipa, Students

I will give an overview of some activities in our institute in order to exemplify the idea of cognitively inspired AI. In particular, I will focus on cooperation projects with IBM, the company that coined the term "cognitive computing". Several terms are currently used for this or similar research endeavors, e.g. human-like AI, human-level AI, AI on a human scale, Artificial General Intelligence, Cognitive Computing, or Cognitive Services. They do not mean the same, but they are somehow related to each other. In my talk, I will briefly sketch applications like a smart city guide, smart farming, and flue prediction that were partially built on IBM's Bluemix services. Additionally I will sketch the idea to build an embodied e-tutor, a project that is currently ongoing. A focus will be to stress aspects that are easy (or straightforward) to implement and aspects that are rather difficult to realize.

3.14 From Turing to Deep Learning: Explaining AI through neurons and symbols

Luis C. Lamb (Federal University of Rio Grande do Sul, BR)

Learning and reasoning have been the subject of great research interest since the dawn of Artificial Intelligence [8]. In the 1950s, Turing already described principles for neural computation, machine learning and formal computational reasoning. Over the last decades, AI research has been questioned and praised on several occasions. However, recent developments in neural learning, in particular deep neural networks, have greatly impacted not only the academic research community, but also have been recognized as a key technology by both the computing industry and popular media channels [7]. In this talk, I presented an overview of the evolution of machine learning in AI, with particular attention to developments and efforts towards integrating (deep) learning and reasoning methods into an unified explainable foundation [2, 5]. I concluded showing that advances in AI, in particular neural-symbolic computation lead to the construction of rich computing systems that integrate neural learning, temporal and cognitive reasoning with applications in several areas [1, 3, 4, 5, 6].

References

1 Rafael V. Borges, Artur d'Avila Garcez, Luis C. Lamb. Learning and Representing Temporal Knowledge in Recurrent Networks. *IEEE Transactions on Neural Networks*, 22(12):2409–2421, Dec. 2011. IEEE Press.
2 A. S. d'Avila Garcez, Marco Gori, Pascal Hitzler, Luís C. Lamb: Neural-Symbolic Learning and Reasoning (Dagstuhl Seminar 14381). Dagstuhl Reports 4(9): 50–84, 2015.
3 A.S. d'Avila Garcez, Luis C. Lamb, and Dov Gabbay. Connectionist Modal Logics: Representing Modalities in Neural Networks. *Theoretical Computer Science*, 371(1–2):34–53, 2007
4 Artur S. d'Avila Garcez, Luis C. Lamb, Dov M. Gabbay. *Neural-Symbolic Cognitive Reasoning*. Springer, 2009.
5 A. d'Avila Garcez, Tarek R. Besold, Luc De Raedt, Peter Földiak, Pascal Hitzler, Thomas Icard, Kai-Uwe Kühnberger, Luis C. Lamb, Risto Miikkulainen, Daniel L. Silver. *Neural-Symbolic Learning and Reasoning: Contributions and Challenges*. Proceedings of the AAAI Spring Symposium on Knowledge Representation and Reasoning: Integrating Symbolic and Neural Approaches. Stanford Univ., March 2015, pp. 18–21, AAAI Press, 2015.
6 H.L. de Penning, Artur S. d'Avila Garcez, Luis C. Lamb, John-Jules Ch. Meyer: *A Neural-Symbolic Cognitive Agent for Online Learning and Reasoning*. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI 2011. Pages 1653–1658, 2011.
7 Y. LeCun, Y. Bengio, G. Hinton. Deep Learning. *Nature*, 521 (7553): 436–444, 2015.
8 Alan Turing. Computing Machinery and Intelligence, *Mind*, LIX (236): 433–460, 1950.

3.15 Interacting Conceptual Spaces

Martha Lewis (University of Oxford, GB)

Joint work of Joe Bolt, Bob Coecke, Fabrizio Genovese, Martha Lewis, Dan Marsden, Robin Piedeleu
Main reference J. Bolt, B. Coecke, F. Genovese, M. Lewis, D. Marsden, R. Piedeleu, "Interacting Conceptual Spaces I : Grammatical Composition of Concepts", arXiv:1703.08314v2 [cs.LO], 2017.
URL https://arxiv.org/abs/1703.08314v2

How should we represent concepts and how can they be composed to form new concepts, phrases and sentences? The categorical compositional distributional programme of [2] successfully integrates two fundamental aspects of language meaning: firstly, the symbolic approach in which meanings of words compose to form larger units; and secondly, the distributional approach where word meanings are derived automatically from text corpora. These two approaches are unified by the key insight that each approach carries the same abstract structure, formalized using category theory.

The abstract framework of the categorical compositional scheme is actually broader in scope than natural language applications. It can be applied in other settings in which we wish to compose meanings in a principled manner, guided by structure. The outline of the general programme is as follows [1]:

1. a. Choose a compositional structure, such as a pregroup or combinatory categorial grammar.
   b. Interpret this structure as a category, the *grammar category*.
2. a. Choose or craft appropriate meaning or concept spaces, such as vector spaces, density matrices, or conceptual spaces.
   b. Organize these spaces into a category, the *semantics category*, with the same abstract structure as the grammar category.
3. Interpret the compositional structure of the grammar category in the semantics category via a functor preserving the necessary structure.
4. This functor then maps type reductions in the grammar category onto algorithms for composing meanings in the semantics category.

In this talk I describe how this programme can be applied to a number of different meaning spaces, including vector spaces, density matrices, conceptual spaces [3], and neural meaning spaces [4] and describe how these fit into the general programme described above. There are also choices for the syntactic side. I give examples of different grammars that can be used, and discuss how grammars might be built based on the semantic category chosen.

Acknowledgements: This work was partially funded by AFSOR grant "Algorithmic and Logical Aspects when Composing Meanings", the FQXi grant "Categorical Compositional Physics", and EPSRC.

References

1 Joe Bolt, Bob Coecke, Fabrizio Genovese, Martha Lewis, Dan Marsden, and Robin Piedeleu. Interacting conceptual spaces i: Grammatical composition of concepts. *arXiv preprint arXiv:1703.08314*, 2017.
2 B. Coecke, M. Sadrzadeh, and S. Clark. Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis*, 36:345–384, 2010.
3 P. Gärdenfors. *Conceptual spaces: The geometry of thought.* The MIT Press, 2004.
4 Paul Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, 46(1-2):159–216, 1990.

3.16 Deep Learning with Symbols

Daniel L. Silver (Acadia University – Wolfville, CA)

Joint work of Daniel L. Silver, Moh. Shameer Igball and Ahmed Galila at Acadia University, Wolfville, NS, Canada (2015-2017)
Main reference M. S. Iqbal, D. Silver, "A Scalable Unsupervised Deep Multimodal Learning System", in Proc. of the 29th Int'l FLAIRS Conf. (FLAIRS 2016), pp. 50–55, AAAI Press, 2016.
URL http://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS16/paper/view/12928
Main reference Y. Bengio, "Evolving culture vs local minima', arXiv:1203.2990v2 [cs.LG], 2012.
URL https://arxiv.org/abs/1203.2990v2
Main reference I. Goodfellow, Y. Bengio, A. Courville, "Deep learning", in Adaptive computation and machine learning, MIT Press, 2016.
URL http://www.deeplearningbook.org/

We present work on the development of (1) a multimodal learning system that uses three sensor/motor channels and a symbolic channel to learn handwritten digits and (2) a system that learns to add handwritten digits aided by symbols of those digits. The conjecture is that symbols are predominately external communication tools that allow agents to share complex noisy concepts that help avoid local minimum in model development. They do so by providing secondary modeling tasks that provide beneficial inductive bias during learning, therefore reducing the number of examples required to accurate learn a new concept. The multimodal learning system uses a visual, auditory and motor channel constructed from stacks of RBMs. These stacks are brought together and connected to one large (2000 neurons) top level RBM and trained using contrastive divergence and a novel back-fitting algorithm. A symbolic channel was also added to display the concept that the activations in the top layer is representing. This addition was found to significantly improve the models performance on all channels. This supports the idea that classification of sensory input as a secondary task provides beneficial inductive bias. In the second network we test this idea by learning to add noisy MNIST digits with and without binary symbolic values at the input and output of a deep BP network. We propose that the models that are augmented with symbols, particularly at the output develop better models. Early experiments support this proposal.

3.17 Grounded Language Processing and Learning

Michael Spranger (Sony CSL – Tokyo, JP)

The talk discusses language learning from the viewpoint developmental robotics. We are interested in investigating how robots can learn to communicate starting from basic sensorimotor capacities through various stages of communicative proficiency: gesture [1], lexicon [3, 4] and grammar [2]. The focus of these experiments is to build grounded models of language processing but also how communication systems can be acquired. We discuss various algorithmic approaches from symbolic to probablistic to neural networks and highlight their current performance and research issues.

References

1 Spranger, Michael and Steels, Luc. *Discovering communication through ontogenetic ritualisation*. Proc. Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob), pages 14–19, IEEE Press, 2014. DOI:10.1109/DEVLRN.2014.6982948.

2 Spranger, Michael and Steels, Luc. *Co-Acquisition of Syntax and Semantics – An Investigation in Spatial Language*. Proc. IJCAI'15: Proceedings of the 24th international joint conference on Artificial intelligence, pages 1909–1915, AAAI Press, 2015.
3 Spranger, Michael and Beuls, Katrien. *Referential Uncertainty and Word Learning in High-dimensional, Continuous Meaning Spaces*. Proc. Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob), IEEE Press, 2016.
4 Nevens, Jens and Spranger, Michael. *Computational Models of Tutor Feedback in Language Acquisition*. Proc. Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob), IEEE Press, 2017.

3.18 Are 'logically naive' people logically naive?

Keith Stenning (University of Edinburgh, GB)

Joint work of Theodora Achourioti, Keith Stenning
Main reference T. Achourioti, A. J. B. Fugard, K. Stenning, "The empirical study of norms is just what we are missing", Frontiers in Psychology, 5:1159, 2014.
URL https://doi.org/10.3389/fpsyg.2014.01159

The categorial syllogism is the oldest topic of the psychology of reasoning with the first paper by Störring [2]. Experimenters have agreed that their subjects are trying to do classical logical reasoning, and that they are, on the whole, doing it very badly. Recently, many have rejected the interest of the syllogism task and argued for probability as the appropriate normative framework for human reasoning. Stenning and Yule [1] showed that it is in fact hard to tell the difference in this task's data between a subject reasoning perfectly by classical logic, and one reasoning with tactically refined nonmonotonic methods; and produced evidence that most subjects can be construed as doing 'preferred model' reasoning in Logic Programming i.e. cooperative nonmonotonic discourse processing.

The multiple logics framework of Stenning & van Lambalgen shows that cooperative expository discourse is well modelled in logic programming, and has been construed as proposing that human reasoning is conducted in LP. Here I will present evidence that 'logically naive' subjects are perfectly capable of trying to reason classically if they are given a task that is clearly intended as the conduct of an adversarial dispute: namely providing counterexamples against invalid inferences proposed by an untrustworthy competitive agent. In short, their counterexamples establish that their understanding of classical conditionals in this situations is material implication, because their understanding of validity in this situation is of the absence of counterexamples. They are by no means tactically adept, but counter modelling in these conditions is a hard task, and induces a completely different attitude than the standard 'draw-a-conclusion' task.

So, at least two logics are needed to model naive human reasoning. This raises crucial questions for the relation between symbolic and sub-symbolic computation. Propositional LP is neurally implementable and highly tractable for searching large knowledgebases: an ideal logic for modelling implicit reasoning (System 1) because its search does not require supervision. Classical logic provided many of the insights into intractable computation, though the tiny syllogistic fragment is tractable. The real questions about the psychological implementation of classical logic are perhaps about the 'theorem prover' that has to guide reasoning, and its implementation.

So, at least two logics are needed to modelling naive human reasoning. This raises crucial questions for the relation between symbolic and sub-symbolic computation.

References

1 Stenning, K. and Yule, P. Image and language in human reasoning: a syllogistic illustration. *Cognitive Psychology*, 34:109–159, 1997.
2 Störring, G. Experimentelle untersuchungen über einfache Schlussprozesse, *Archiv für die Gesamte Psychologie*, 11:1–127, 1908.

3.19 Logic Programming as a framework for understanding human computation

Keith Stenning (University of Edinburgh, GB)

Main reference K. Stenning, M. van Lambalgen, "Human reasoning and Cognitive Science", MIT Press, 2008.
URL http://mitpress.mit.edu/books/human-reasoning-and-cognitive-science

Logic Programming (LP) has been used for analysing discourse and human reasoning with some success. The propositional form of LP is neurally implementable [1, 2]. This talk will explore the intersection this overlap between AI, Computer and Cognitive Science offers, raising some issues, from the cognitive side of the fence.

I will introduce four of the many issues that arise, chosen for possible relevance to neural-symbolic computation: 1) Computational tractability and questions of scaling-up; 2) Distinguishing reasoning and learning in psychology and computer science; 3) The relation of extensional to intensional reasoning; and 4) Qualitatively distinct kinds of uncertainty (especially relations between analyses in LP and probability). If discussion does not detain us, the talk will finish with an example of a current project at the nitty-gritty level: an analysis of human semantic memory in the service of nonmonotonic reasoning.

References

1 M. van Lambalgen, F. Hamm, *The Proper Treatment of Events*. Blackwell, Oxford and Boston, 2004.
2 Stenning, K. and van Lambalgen, M. *Human reasoning and Cognitive Science*. MIT Press, Cambridge, MA, 2008.

3.20 Structured Computer Organization of the Human Mind

Niels A. Taatgen (University of Groningen, NL)

Main reference N. A. Taatgen, "The nature and transfer of cognitive skills", Psychological Review, 120(3):439–471, 2013.
URL http://psycnet.apa.org/doi/10.1037/a0033138

Cognitive architectures aim to bridge the gap between the brain and behavior, providing a formal level of description that provides a basis for rigorous theories of behavior. Architectures typically operate on a single level of abstraction. In my talk I argued that we need multiple levels of abstraction, each with their own formalisms and learning mechanisms. Each of these levels should be to explain the abstraction level above, creating a reductionist hierarchy of theories that can bridge the gap, not with a single formalism, but with several.

3.21 Embodied neuro-symbolic computation

Serge Thill (University of Skövde, SE)

This talk covers two distinct reasons I am interested in neuro-symbolic integration. The first is because one is creating a cognitive system acting in the real world and carrying out some non-trivial task. In such systems, one often encounters problems that are better addressed with subsymbolic solutions such as neural networks and others for which symbolic approaches are more amenable. Here, I present the new European project Dreams4Cars (Grant agreement nr 731593) as one example of such a case.

The second reason is because it is a way to test representationalist embodied models of cognition: if embodied theories are right, then the sensorimotor experience underlying the formation of human concepts is critical for those concepts, and fundamentally shapes the computations that use them. An example effect one might expect in such a case is that the differences in sensorimotor experience between two agents might affect the degree to which they can communicate about concepts, or operations possible on concepts. It is therefore interesting to build a model of symbol grounding in a "rich" sensorimotor experience, and this cannot ignore the subsymbolic neural levels (since neural computations are fundamentally shaped by the constraints of biology, and possibly the specific needs of the organism) nor the details of the human sensorimotor experience (which, for example, includes interoceptive features).

The present talk attempts an overview of interesting issues in both directions.

3.22 Overview of neuro-symbolic processing in Neural Blackboard Architectures

Frank Van der Velde (University of Twente)

Joint work of Frank Van der Velde, Marc de Kamps
Main reference F. van der Velde, M. de Kamps, "Neural blackboard architectures of combinatorial structures in cognition", Behavioral and Brain Sciences, 29:37–70, 2006.
URL https://doi.org/10.1017/S0140525X06009022

Neural Blackboard Architectures (NBAs) aim to account for and simulate combinatorial structures and processing in a neuronal manner. Examples of these include sentential structures and reasoning with relations, but NBAs also include perceptual processing (vision). The focus in this presentation will be on sentential structures, both in relation to human processing and as a computational architecture for parallel processing. NBAs satisfy a number of constraints we assume to exist in neural and (human) cognitive processing. These include the assumption that representations (e.g. of "words") are grounded 'in situ' representations, also in combinatorial structures. NBAs provide "connection paths" between such representations, as a basis for generating behavior. Representations are always content-addressable in the underlying connection structures, also when they are a part of combinatorial structures such as sentences. Simulations with NBAs will be illustrated with incremental sentence processing and dynamical competitions, which can account for ambiguity resolution and garden path effects as found in human behavior.

References

1. van der Velde, F. and de Kamps, M. *From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation*. Journal of Cognitive Neuroscience, 13(4), 479–491, 2001.
2. van der Velde, F. and de Kamps, M. *Learning of control in a neural architecture of grounded language processing*. Cognitive Systems Research, 11, 93–107, 2010.
3. van der Velde, F. and de Kamps, M. *Compositional connectionist structures based on in situ grounded representations*. Connection Science, 23, 97–107, 2011.
4. van der Velde, F. *Communication, concepts and grounding*. Neural Networks, 62, 112–117, 2015.
5. van der Velde, F. and de Kamps, M. *The necessity of connection structures in neural models of variable binding*. Cognitive Neurodynamics, 9, 359–370, 2015.
6. van der Velde, F. and de Kamps, M. *Combinatorial structures and processing in neural blackboard architectures*. In Proceedings of the Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches (CoCo@NIPS 2015), eds T. R. Besold, A. d'Avila Garcez, G. F. Marcus, and R. Miikkulainen (Montreal), (pp. 1–9). CEUR Workshop Proceedings, Vol. 1583, 2015.
7. van der Velde, F. *Concepts and relations in neurally inspired in situ concept-based computing*. Frontiers in Neurorobotics. 10:4, 2016.

3.23 Neural Computing with Signals and Symbols for Music Analysis

Tillman Weyde (City, University of London, GB)

Understanding music with computational methods is a challenging but rewarding task, as music is an essential and universal feature of humanity. Music information retrieval research develops methods that extract information from music data at different levels, increasingly using neural networks. The different representation levels of music as audio signal, notes, harmony, form etc, as well as emotional and cultural aspects, naturally create a need to connect unstructured and symbolic representations. In this talk, examples are discussed to demonstrate trends, achievements, and challenges in the development of human-like music computation.

References

1. Andreas Jansson and Eric Humphrey and Nicola Montecchio and Rachel Bittner and Aparna Kumar and Tillman Weyde. *Singing Voice Separation with Deep U-Net Convolutional Networks*. In Proc. of the International Society for Music Information Retrieval, ISMIR 2017, Suzhou, China, 2017 (in press).
2. Siddharth Sigtia and Emmanouil Benetos Nicolas Boulanger-Lewandowski and Tillman Weyde and Artur d'Avila Garcez and Simon Dixon. *A hybrid recurrent neural network for music transcription*. In Proc. of the International Conference on Acoustics, Speech and Signal Processing, ICASSP, Brisbane, Australia, 2015.

3 Reinier de Valk and Tillman Weyde. *Bringing 'Musicque into the tableture': machine-learning models for polyphonic transcription of 16th-century lute tablature*. Early Music 43.4 (2015): 563–576.
4 Tilman Weyde and Klaus Dalinghaus. *A neuro-fuzzy system for sequence alignment on two levels*. Mathware & Soft Computing 11.3 (2004):197–210.

4 Working groups

4.1 Logic Tensor Networks Hackathon

Lucas Bechberger (Universität Osnabrück, DE), Artur d'Avila Garcez (City, University of London, GB), Raquel Garrido Alhama (University of Amsterdam, NL), Marco Gori (University of Siena, IT), Luciano Serafini (Bruno Kessler Foundation – Trento, IT), Michael Spranger (Sony CSL – Tokyo, JP), and Tillman Weyde (City, University of London, GB)

Logic Tensor Networks (LTNs) [1] provide a neural network based solution for combining symbolic knowledge (encoded as first-order logic relations) and subsymbolic knowledge (represented as feature vectors). Each object is described as a vector of real numbers which together with relational symbolic knowledge get translated into soft and hard constraints on the subsymbolic level (implemented as a tensor network). The network then learns to approximate a solution to the constraint-optimization problem called best satisfiability when faced with new data. LTNs enable therefore relational knowledge infusion into deep networks, and knowledge completion and distilling through querying the networks.

In this hackathon, we investigated the application of the LTN framework to the domain of word embeddings. A word embedding such as word2vec [2] supplies for each word in language a so called "embedding", i.e. a vector in a high-dimensional space. These vectors are usually learned based on co-occurence patterns in a large text corpus. Somewhat surprisingly, the vectors of semantically related words tend to have a relatively low cosine distance. One can even go further and perform some arithmetic on the vectors (the vector resulting from "king – man + woman" is quite similar to the vector for "queen").

In our application scenario, we tried to generate new word vectors for previously unknown words based on some logical constraints that could for instance come from a dictionary entry or from an ontology. Our showcase looked as follows: we assumed that there is no vector embedding for the word "zebra", but there are embeddings for "horse" and for "black and white stripes". By defining an object as a zebra if and only if it is a horse and has black and white stripes, we wanted to derive a word embedding for "zebra". For a first proof-of-concept implementation, we used only a two-dimensional space with the following vectors for "horse" and "black and white": horse = (1.0, 0.0), black and white = (0.0, 1.0). We also added vectors for some additional unrelated words (e.g. cow). When running the LTN framework multiple times, we typically obtained vectors in the first quadrant, i.e. vectors for which both coordinate entires were greater than zero, as should be expected.

This was found to serve as a good example for introducing LTNs to academics with distinct backgrounds (e.g. symbolic logic/AI and statistical Machine Learning) and to help them get familiarized with the Tensorflow implementation of LTN, which is available at https://github.com/LucianoSerafini/dagstuhl_hackaton_LTN.

Following results reported in [1] of the application of LTNs to semantic image interpretation (c.f. https://uk.arxiv.org/abs/1705.08968) which can improve the performance of state-of-the-art image detection with convolutional networks, we are optimistic that LTNs can also help improve results in the natural language processing area by connecting word embeddings to symbolic knowledge from ontologies.

References

1. Ivan Donadello, Luciano Serafini, and Artur d'Avila Garcez. *Logic Tensor Networks for Semantic Image Interpretation*. Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017
2. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. *Distributed Representations of Words and Phrases and Their Compositionality*. Advances in Neural Information Processing Systems, 2013.

4.2 Explainability Hackathon

Tarek R. Besold (Universität Bremen, DE), Derek Doran (Wright State University – Dayton, US), Pascal Hitzler (Wright State University – Dayton, US), Caroline Jay (University of Manchester, GB), Oliver Kutz (Free University of Bozen-Bolzano, IT), and Sarah Schulz (Universität Stuttgart, DE)

Quite frequently demands for better comprehensible and explainable Artificial Intelligence (AI) and Machine Learning (ML) systems are being put forward. The explainability hackathon was intended to shed light on notions such as "comprehensibility" and "explanation" in the context of AI and ML, working towards a better understanding of what an explanation is when talking about intelligent systems, what it means to comprehend a system and its behavior, and how human-machine interaction can take these dimensions into account.

The hackathon compared notions of comprehension and explanation as used in different fields, ranging from philosophy to cognitive science, and mapped the findings back into AI and ML, aiming to establish generic types of systems. Three conceptual categories emerged, namely systems offering no insight into their input/output mechanisms; systems where users can mathematically analyze these mechanisms; and systems emitting symbols that allow a user to draw an intuitive explanation of how a conclusion is reached.

Following the Dagstuhl seminar, this initial line of work has been further expanded into a position paper titled "What Does Explainable AI Really Mean? A New Conceptualization of Perspectives" which has been submitted to the AI*IA 2017 workshop on Comprehensibility and Explanation in AI and ML (CEx). In that paper, additionally efforts towards systems of a fourth type are encouraged: explainable systems, where automated reasoning is central, yet missing from much of the current work.

4.3 Deep Learning with Symbols Hackathon

Daniel L. Silver (Acadia University – Wolfville, CA), James Christopher Davidson (Google Inc. – Mountain View, US), Dieuwke Hupkes (University of Amsterdam, NL), Isaac Noble (Playground Global Inc., US), and Katja Seeliger (Radboud University Nijmegen, NL)

Background: Recent work by Igbal and Silver (FLAIRS-2016 Best Paper Award) [1] inspired by work by Hinton et al. [2] and Srivastava et al. [3] has shown that it is possible to develop a multimodal deep learning system for learning a noisy handwritten digits using four sensor/motor channels (visual, audio, robotic, and symbolic) and an associative layer that ties all channels together. After training, the presentation of a digit (sound, image, drawing) at the visible nodes of the model activates all other channels to create their associated reconstruction at their respective visible nodes. Each channel provides additional information that helps the other channels more accurately reconstruct the output at their visible nodes. The symbolic channel outputs the cleanest and clearest signal as to what digit the multimodal deep learning system is "thinking" of given input on another channel. The symbolic channel also provides the cleanest and clearest input to assist other channels to generate the correct reconstructions at their visible nodes. This led us to a paper by Yoshua Bengio [4] that discusses the value of symbols (ie. language) in helping individuals to learn concepts (like "cat") better without having to see all possible examples of that concept. This has a profound impact upon the development of our culture and the human species.

Purpose: To create a deep recurrent neural network architecture that can add two or more noise MNIST digits in a row and produce the correct symbolic response at the output. The background provided above inspired us to consider a learning agent that learns to perform a mathematic operation using two noising channels but which can (at times) also receive concise information on a symbolic channel about the data on the noisy channels. We chose to investigate a system that could learn to add a sequence of handwritten digits aided (on occasion) by symbols of those digits at additional inputs to the same network. The conjecture is that symbols are predominately external communication tools that allow agents to share complex noisy concepts that help avoid local minimum in model development. They do so by providing secondary tasks that provide beneficial inductive bias during learning, therefore reducing the number of examples required to accurate learn a new concept. Our intention was to test this idea by developing a deep recurrent neural network using Keras/Tensorflow to add a sequence of noisy MNIST digits with and without binary symbolic values at the inputs and outputs of the network.

Progress: With much coding work by Dieuwke and Katja and support from Danny, James, and Isaac we were able to develop a Keras/Tensorflow-based LSTM recurrent network that could accept two sequential MNIST digits and output a symbolic value (two sets of 10 binary nodes) indicating the estimated result of the addition. Some of us intend to pursue this direction further using similar methods.

References

1 M.S. Iqbal and D. Silver. A Scalable Unsupervised Deep Multimodal Learning System. Proceedings of the Twenty-Ninth International FLAIRS Conference (FLAIRS-29), Key Largo, FL, AAAI Press, pp.50-55, 2016.
2 G.E. Hinton, S. Osindero and Y.-W. Teh. A fast learning algorithm for deep belief nets. Neural computation 18(7):1527– 1554, 2016.

3 N. Srivastava and R. R. Salakhutdinov. Multimodal learning with deep Boltzmann machines. In Advances in neural information processing systems, 2222–2230, 2012.
4 Yoshua Bengio. Evolving Culture vs Local Minima, March, 2012, https://arxiv.org/abs/1203.2990

4.4 Nengo/ACT-R/PRIMs/Cognitive Architectures Hackathon

Niels A. Taatgen (University of Groningen, NL), Jelmer Borst (University of Groningen, NL), Marc de Kamps (University of Leeds, GB), Martha Lewis (University of Oxford, GB), Serge Thill (University of Skövde, SE), and Frank Van der Velde (University of Twente)

Cognitive architectures aim to explain and simulate various or all aspects of human cognition and intelligence. Some of these architectures model cognition using symbolic representations, or representations close to symbolic, whereas others focus on neural simulations. A challenge for future progress is to integrate both approaches in order to benefit from accomplishments of both. In this Hackathon we discussed this integration between ACT-R and PRIMs (symbolic) on the one hand, and Nengo and the Neural Blackboard architecture on the other hand. In particular, we discussed how the smallest elements of skill proposed by the PRIMs architecture (primitive operations) can be implemented in a Basal Ganglia model in either Nengo or the Neural Blackboard architecture. As benchmark to test such models, we selected a set of monkey studies in which the monkey had to make controlled eye-movements. In addition, we talked about how conceptual spaces models can be viewed in the light of cognitive/neural architectures. To further develop this research, we agreed to organize a future workshop in Amsterdam in December 2017.

5 Open problems

5.1 Research Challenges

Luis C. Lamb (Federal University of Rio Grande do Sul, BR), Tarek R. Besold (Universität Bremen, DE), and Artur d'Avila Garcez (City, University of London, GB)

Recent developments in AI, in particular machine learning and deep networks demands explainability. In particular, humans need to trust AI systems and applications. Novel, state-of-the-art AI systems will have to provide explanations for their actions, behaviours and their impact on humanity. As mentioned above, this seminar consisted of both invited and contributed talks, joint group discussion sessions, and scientific hackathons. Several questions and research challenges were identified. One clear point made by the participants is that Human-Like Neural-Symbolic Computing is an effective methodology towards explainable AI [2]. As a result, Neural-Symbolic Computing clearly contributes towards answering questions that are both scientifically, ethically and methodologically relevant in the context of current Artificial Intelligence research.

In summary, the main research directions identified in the seminar are:

- The recent success of deep learning in computer vision and language processing [10], associated with the growing complexity of big data applications demands the development of *explainable AI models*. Neural-symbolic computation offers several methods and principles that contribute towards this aim. For instance, rule extraction mechanisms [3] and the integrated methodology for the development of neural-symbolic cognitive systems offers principled alternatives [1, 5, 6, 8, 9]. The concept of modularity in deep networks can be explained by using neural-symbolic methods. Modularity is suitable to knowledge extraction (as identified in [4, 5]) which can result in the development of effective knowledge extraction methods towards explainable AI.
- The development of Cognitive Architectures capable of simulating and explaining aspects of human cognition also remains an important research endeavour [11]. Some cognitive architectures typically consider symbolic representations, whereas others employ neural simulations. The integration of these models remains a challenge and there are benefits on integrating the accomplishments of both paradigms, as identified in the cognitive architectures Hackathon.
- Logic Tensor Networks (LTNs) [7] provide a model that integrates symbolic knowledge (encoded as first-order logic relations) and subsymbolic knowledge (represented as feature vectors). LTNs enable the representation of relational knowledge infusion into deep networks, and knowledge completion and distilling through querying the networks. There remains a number of challenges in integrating, explaining and computing symbolic knowledge in deep networks. Both LTNs [7] and Connectionist Modal and Temporal Logics [5, 6, 9] offer effective alternatives towards these research challenges.

References

1. Tarek R. Besold, Artur S. d'Avila Garcez, Keith Stenning, Leendert W. N. van der Torre, Michiel van Lambalgen: Reasoning in Non-probabilistic Uncertainty: Logic Programming and Neural-Symbolic Computing as Examples. *Minds and Machines*, 27(1): 37–77, 2017.
2. A. d'Avila Garcez, Tarek R. Besold, Luc De Raedt, Peter Földiak, Pascal Hitzler, Thomas Icard, Kai-Uwe Kühnberger, Luis C. Lamb, Risto Miikkulainen, Daniel L. Silver. Neural-Symbolic Learning and Reasoning: Contributions and Challenges. Proceedings of the AAAI Spring Symposium on Knowledge Representation and Reasoning: Integrating Symbolic and Neural Approaches. Stanford Univ., March 2015, pp. 18–21, AAAI press, 2015.
3. Artur S. d'Avila Garcez, Dov M. Gabbay, Krysia Broda. *Neural-Symbolic Learning System: Foundations and Applications*. Springer-Verlag, New York, Inc., USA, 2002.
4. A. S. d'Avila Garcez, Marco Gori, Pascal Hitzler, Luís C. Lamb: Neural-Symbolic Learning and Reasoning (Dagstuhl Seminar 14381). Dagstuhl Reports 4(9): 50–84, 2015.
5. Artur S. d'Avila Garcez and Luis C. Lamb. A Connectionist Computational Model for Epistemic and Temporal Reasoning. *Neural Computation*, 18(7):1711–1738, 2006.
6. Artur S. d'Avila Garcez, Luis C. Lamb and Dov M. Gabbay. *Neural-Symbolic Cognitive Reasoning*. Springer Publishing Company, 2009. ISBN: 9783540732457.
7. Ivan Donadello, Luciano Serafini, and Artur d'Avila Garcez. *Logic Tensor Networks for Semantic Image Interpretation*. Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017.
8. Barbara Hammer, Pascal Hitzler (Eds.): Perspectives of Neural-Symbolic Integration. Studies in Computational Intelligence 77, Springer 2007.
9. Luis C. Lamb, R.V. Borges and A.S. d'Avila Garcez. *A connectionist cognitive model for temporal synchronisation and learning*. Proc. AAAI Conference on Artificial Intelligence AAAI-07, pages 827–832, AAAI Press, 2007.
10. Y. LeCun, Y. Bengio, G. Hinton. Deep Learning. *Nature*, 521 (7553): 436–444, 2015.
11. K. Stenning and M. van Lambalgen. *Human reasoning and Cognitive Science*. MIT Press, Cambridge, MA, 2008.

Participants

- Lucas Bechberger
Universität Osnabrück, DE
- Tarek R. Besold
Universität Bremen, DE
- Jelmer Borst
University of Groningen, NL
- Artur d'Avila Garcez
City, University of London, GB
- James Christopher Davidson
Google Inc. –
Mountain View, US
- Marc de Kamps
University of Leeds, GB
- Derek Doran
Wright State University –
Dayton, US
- Ulrich Furbach
Universität Koblenz-Landau, DE
- Raquel Garrido Alhama
University of Amsterdam, NL
- Marco Gori
University of Siena, IT
- Pascal Hitzler
Wright State University –
Dayton, US
- Dieuwke Hupkes
University of Amsterdam, NL
- Caroline Jay
University of Manchester, GB
- Kristian Kersting
TU Darmstadt, DE
- Kai-Uwe Kühnberger
Universität Osnabrück, DE
- Oliver Kutz
Free University of
Bozen-Bolzano, IT
- Luis C. Lamb
Federal University of
Rio Grande do Sul, BR
- Martha Lewis
University of Oxford, GB
- Isaac Noble
Playground Global Inc., US
- Sarah Schulz
Universität Stuttgart, DE
- Katja Seeliger
Radboud University
Nijmegen, NL
- Luciano Serafini
Bruno Kessler Foundation –
Trento, IT
- Daniel L. Silver
Acadia University –
Wolfville, CA
- Michael Spranger
Sony CSL – Tokyo, JP
- Keith Stenning
University of Edinburgh, GB
- Niels A. Taatgen
University of Groningen, NL
- Serge Thill
University of Skövde, SE
- Frank Van der Velde
University of Twente, NL
- Tillman Weyde
City, University of London, GB