Data, Knowledge, and Computation

Christian Igel¹

Accepted: 5 August 2021 / Published online: 11 August 2021
© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Dear readers,

A journalist asked to comment on Richard Sutton’s 2019 essay “The Bitter Lesson”, published online at <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>. It is worth reading, and I would like to share my thoughts about the essay with you in this editorial.

Richard S. Sutton is a true AI pioneer, known for his seminal work on Reinforcement Learning [e.g. 2, 11, 12]. Verena Heidrich-Meisner interviewed him for KI in 2009 [6]. In “The Bitter Lesson”, Sutton concludes:

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin.

He gives search and learning as prime examples of such general methods. Referring to natural language and speech processing, AI for games, and computer vision, he argues that “researchers always tried to make systems that worked the way the researchers thought their own minds worked—they

tried to put that knowledge in their systems—but it proved ultimately counterproductive”. He bases his claim on the undeniable success of deep learning systems exploiting huge amounts of training data and compute resources.

Not surprisingly, as a machine learning researcher I agree in principle with the general praise of search and learning, but let me add some thoughts on the role of prior knowledge in AI and upscaling as a general solution strategy.

Learning that is “leveraging computation” has to go hand in hand with the availability of huge training data sets. Ultimately, data availability limits the performance even in the case of otherwise infinite compute resources. The way I read the essay, it assumes that scaling of learning systems entails increasing data availability. Let us briefly look at this situation from the theoretical perspective. In general, when the amount of (i.i.d.) data increases, distributions of statistical estimates become more concentrated. There are many learning algorithms that converge to the best possible solution when the number of training data points goes to infinity [5, 9]. However, for estimates based on finite data we have *no-free-lunch*: For any algorithm there is a data distribution for which it works arbitrarily bad given a finite sample [3, 5]. If “there is no a priori restriction on the possible phenomena that are expected, it is impossible to generalize and there is thus no better algorithm (any algorithm would be beaten by another one on some phenomenon)” [3]. As Bousquet et al. nicely put it:

Generalization = Data + Knowledge

This knowledge about possible phenomena can be rather general (e.g., some weak notion of continuity that can be assumed at least partly because the problem at hand is rooted in the physical world) and may already be reflected by the choice of a rather general learning algorithm.

All systems underlying the success stories of learning and optimization mentioned in Sutton’s essay still encode information about the problem at hand in some way, but not necessary about the way humans solve this problem. State-of-the-art learning systems for images, speech, and natural language use prior knowledge about the spatial and temporal

✉ Christian Igel
igel@di.ku.dk

¹ Department of Computer Science (DIKU), University of Copenhagen, 2100 Copenhagen Ø, Denmark

structure of the data. For example, convolutional and pooling operations are used because of a priori assumptions about the spatiotemporal structure of the input; recurrent neural networks are chosen to exploit the sequential nature of data. This use of knowledge about the problem is also appreciated in Sutton's essay ("deep-learning neural networks use only the notions of convolution and certain kinds of invariances"). Data augmentation, which is a common technique to improve generalization when the data is limited, also generally relies on prior knowledge about invariance properties. Incorporation of known constraints is also helpful. For example, the recently published seminal deep learning system AlphaFold [8] for determining the 3D shapes of proteins incorporates information about the physical and geometric constraints that govern protein folding. In conclusion, for efficiently and effectively solving machine learning tasks we have to bring in knowledge about the problem, but not necessary about how humans solve it.

In my opinion, the essay lacks an acknowledgement that in practice scaling up may often not be an option and resource efficient solutions are needed. A difficult AI problem may require huge amounts of data and processing time. For some AI applications, data may be very limited, for example for privacy or economical reasons, and there may be not time to wait until this situation changes (if it can change at all). Incorporation of knowledge may compensate for the lack of data and computational resources. Last but not least, the energy consumption of large AI systems (the state-of-the-art natural language processing tool GPT-3 has 175 billion parameters [4]) is already problematic from an environmental point of view [7, 10], which questions scaling up compute systems as a general approach. I would like to use this opportunity to promote Carbontracker [1], a tool for measuring the energy consumption of your deep neural networks and other machine learning models.

Best wishes and enjoy reading this issue of KI,

Christian Igel

Forthcoming Special Issues

Learning Computational Thinking

Guest Editors: Nina Bonderup Dohn (University of Southern Denmark), Yasmin Kafai (University of Pennsylvania),

Anders Mørch (University of Oslo), and Marco Ragni (University of Southern Denmark)

Scope: Our call is informed by the ever-increasing digitalization and integration of AI into everyday life. Computational thinking (CT) is the use of problem-solving processes to enable humans to design creative and computational solutions to many of the wicked problems which the world faces today. Utilizing computational thinking is an integral part of developing AI systems – and the AI developers of tomorrow are the learners of today. CT is not restricted to algorithmic thinking, computer visualizations, and programming, but includes as well analogue representations and even aspects of embodied cognition. We call for papers that develop cognitive, educational, and computational models, report psychological studies, in situ case studies of design activities, children's engagement in computational participation and critical reflection on IT and AI, and investigate pedagogical designs that support the learning of CT and AI using a range of different tools and methods, e.g., computer visualizations, simulation programs, analogue algorithmic thinking, bodily interactions, or physical things that can be programmed, or using all methods from AI. We aim to integrate researchers from all disciplines to present novel research to advance the understanding of CT. All submissions will be peer-reviewed.

The topics of interest for the special issue include, but are not limited to:

- Theoretical investigations of CT compared with other types of thinking.
- Any form of AI support for AI thinking and developing AI thinking.
- Analysis of AI models for taxonomy of AI thinking abilities.
- Comparison of analogue and digital forms of CT.
- Multimodal interactions of users and systems.
- Comparison of successful educational methods for AI and CT (Fachdidaktik).
- In situ experiments with learning designs supporting the learning of CT and AI.
- Practice- or design-based research on CT in different disciplines.
- Case studies of learners engaged in CT activities, computational participation, and critical reflection on AI in formal and informal learning.
- AI interfaces or Apps for enabling learners to engage in AI and CT.

Explainability in Machine Learning

Guest Editors: Barbara Hammer, Andreas Holzinger, Benjamin Paassen, Wojciech Samek

Scope: Machine learning (ML) technologies—in particular deep learning architectures—have led to recent breakthroughs

in various domains and today, ML components are used in a number of applications which severely impact our daily life such as recommendation systems, medical decision support, assisted driving, or video surveillance. Yet in particular deep learning is considered a black box technique and quite some ML models can display unexpected or unwanted behavior caused by adversarial attacks or biased data, for example. This might be a severe threat in safety-critical applications, it can considerably reduce trust, and it poses quite some challenges on how to solve the legal issue of accountability in case of failures. Therefore, a many approaches aim for supplementing, extending or substituting current black box methodologies to make machine learning methods more interpretable, whereby also a definition of the term “explainability” and efficient methods how to evaluate whether a model is interpretable is yet a matter of debate. The aim of this special issue is to provide a compact overview of this growing field. We invite contributions from researchers and practitioners interested in methods, concepts, and applications in the field of explainable machine learning. If you are interested in submitting a paper, please contact one of the guest editors.

Contact: Barbara Hammer
bhammer@techfak.uni-bielefeld.de

Application of AI in Digital Forensics

Guest Editors: Johannes Fährndrich, Roman Povalej, Heiko Rittelmeier, Silvio Berner

Scope: With the increase of digitalization and the pervasiveness of information systems, a crime scene is no longer what it used to be with its mix of a location, people, evidence, changes in time, and their virtual counterpart. Including the mainstream use of smart-homes, -infrastructure, -factories, or -cities, investigations and forensic evidence are no longer bound by physics. With the growing amount of digital information, an application of Artificial Intelligence (AI) in forensics is incumbent. Methods from Machine Learning and Data Science need to be extended to be explainable and valid for legal purposes. This special issue has the goal of collecting work on AI with the application on forensic science with the focus on the amalgamation of computer science, data analytics, and machine learning with the discussion of the law and ethics for its application to cyber forensics. Topic might be, but are not restricted to:

- NLU/NLP in forensic evidence.
- Explainable AI which can stand up in court.
- AI and Object Detection.
- AI and Super resolution.

- AI and Darknets and hidden services investigation.
- AI and emotion recognition.
- AI and lie detection.
- AI and Cybercrime related investigations.
- Fooling neuronal networks and other Anti-forensic techniques and methods Automated analysis for forensic evidence in IoT.
- AI in incident response, investigation and evidence handling.
- Ethical, legal, and societal challenges of using AI in digital forensics

Contact: Johannes Fährndrich
johannesfaehndrich@hfpol-bw.de

References

1. Anthony L, Kanding B, Selvan R (2020) Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. In: ICML workshop on challenges in deploying and monitoring machine learning systems. <https://github.com/lfwa/carbontracker>
2. Barto AG, Sutton RS, Anderson CW (1983) Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans Syst Man Cybern SMC* 13(5):834–846
3. Bousquet O, Boucheron S, Lugosi G (2004) Introduction to statistical learning theory. In: Bousquet O, von Luxburg U, Rätsch G (eds) Summer School on Machine Learning 2003, vol 3176. Springer, LNAI, New York, pp 169–207
4. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A et al (2020) Language models are few-shot learners. *Advances in Neural Information Processing Systems (NeurIPS)*. [arXiv:2005.14165](https://arxiv.org/abs/2005.14165)
5. Devroye L, Györfi L, Lugosi G (2013) A probabilistic theory of pattern recognition. Springer, New York
6. Heidrich-Meisner V (2009) Interview with Richard S. Sutton *Künstliche Intell* 23(3):41–43
7. Henderson P, Hu J, Romoff J, Brunskill E, Jurafsky D, Pineau J (2020) Towards the systematic reporting of the energy and carbon footprints of machine learning. *J Mach Learn Res* 21(248):1–43
8. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Židek A, Nelson AW, Bridgland A et al (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577(7792):706–710
9. Steinwart I (2005) Consistency of support vector machines and other regularized kernel classifiers. *IEEE Trans Inf Theory* 51(1):128–142
10. Strubell E, Ganesh A, McCallum A (2019) Energy and policy considerations for deep learning in NLP. In: Proceedings of the 57th annual meeting of the association for computational linguistics (ACL), pp 3645–3650
11. Sutton RS (1988) Learning to predict by the methods of temporal differences. *Mach Learn* 3(1):9–44
12. Sutton RS, McAllester D, Singh S, Mansour Y (2000) Policy gradient methods for reinforcement learning with function approximation. In: Advances in neural information processing systems, pp 1057–1063