KI - Künstliche Intelligenz (2020) 34:127–130
https://doi.org/10.1007/s13218-020-00662-x

EDITORIAL

Challenges in Interactive Machine Learning

Toward Combining Learning, Teaching, and Understanding

Stefano Teso[1] · **Oliver Hinz**[2]

1 Why Interactive Machine Learning?

The recent success of AI, and of machine learning in particular [8], has unlocked a number of high-impact applications, from machine translation to medical diagnosis, to image and video generation, to game playing. The bulk of these success stories, however, take place in relatively "lab" settings [3, 11]. With this we mean settings in which the cost of failure is limited and mistakes are easily spotted, large repositories of data are available and more supervision can be acquired relatively cheaply, noise patterns are somewhat predictable, the rules of the game and the knowledge they rely on are crisply defined, and the objective function is fully specified from the get-go [6, 10].

Most importantly, many of these success stories take place in settings — like passive learning — that are completely devoid of people [1].[1] This is not insignificant, because once humans are added to the equation, all of the simplifying assumptions listed above break down. As AIs become ever more ubiquitous, the question becomes: will they work appropriately once deployed in the real world, in which humans feature so prominently? It seems prudent to prepare for a negative answer by investigating appropriate human-aware approaches.

✉ Stefano Teso
stefano.teso@unitn.it

Oliver Hinz
ohinz@wiwi.uni-frankfurt.de

[1] University of Trento, Trento, Italy

[2] Goethe University Frankfurt, Frankfurt, Germany

2 About this Special Issue

This special issue is dedicated to interactive machine learning, in which the goal is precisely to design adaptive agents that support meaningful and beneficial interaction with humans.

The article by Nadj et al. identifies and discusses design principles for interactive labeling systems by conducting a literature review. We believe that this contribution can represent a helpful starting point for further efforts to refine and expand the design of interactive labeling systems.

The history of interactive learning is rooted in query learning [2]. Here, the structure of the interaction is fixed: a student model, usually a classifier, acquires supervision by asking questions to one or more human teachers, usually domain experts. This helps explaining why the most studied — and thus the most well-known — issues in interactive learning are related to acquiring supervision, e.g., balancing between the cost of annotations and the amount of information they carry, as in active learning [14], or measuring and improving the quality of annotations and annotators, as in crowd-sourcing applications [5]. The range of issues spanned by interactive machine learning, however, is even broader. The articles in this issue constitute a representative and diverse sampler.

Including humans in the process, certainly comes with higher costs for manual labor. Therefore it is imperative to offer efficient processes and systems that minimize this part of the entire interactive machine learning process. Baur et al. make a significant contribution to solve this problem. They introduce a workflow that enables an "explainable cooperative machine learning" and provide evidence for its superiority by offering a data annotation and model training tool called NOVA. NOVA offers a collaborative annotation

[1] Notable exceptions, like recommender systems [13, 15] and preference elicitation [4, 12], fall within the broader range of interactive machine learning.

backend where multiple annotators can work on the same task. This system allows semi-supervised active learning techniques already during the annotation process by giving the possibility to pre-label data automatically. This makes NOVA a tool that substantially accelerates the annotation process.

The work of Brust et al. generalizes active learning to incremental settings, in order to support the continuous exploration of newly observed unlabeled datapoints. This novel setup shows promising results on object detection and – together with a weakly supervised system – on a real-world biodiversity task.

Kumaraswamy et al. show that user intervention plays a role in aiding transfer learning algorithms beyond the standard classification setting. They tackle "deep" cross-domain transfer, in which a human expert and a transfer learning algorithm collaborate to transfer knowledge across seemingly unrelated domains. The proposed system makes use of rich (first-order) representations to facilitate domain independence. A key aspect of this work is that the expert's guidance is not in the form of labels: the expert helps the machine by prioritizing certain branches of the search tree or by editing and improving the knowledge used by the algorithm. The system includes a novel interface to support this novel form of interactive transfer learning.

Two other contributions focus on the role of computational explanations, such as those sought in the field of explainable AI [7], and to their application in the context of human-machine interaction. The overarching theme is that explanations can serve as an antidote to the progressive sophistication and opacification of models learned from data, especially of high-performance ones, and can facilitate controlling such models [16]. Holzinger et al. tackle the notorious lack of a standard measures for evaluating explanation quality. This stands out as a roadblock for the development of truly understandable computational explanations and interaction protocols. Holzinger et al. address this issue by introducing the System Causability Scale, which combines causability and usability.

Finally, the work of Abdel-Karim et al. explores an understudied aspect of interactive machine learning: the effect of the interaction on the human-in-the-loop as a student. This subverts the notion that the annotator is an immutable oracle. A study on lung X-ray classification shows that human experts reconsider their initial assessment upon observing cases in which a predictor contradicts their opinions. This opens the door to novel collaboration strategies between decision makers and intelligent adaptive systems. Especially under the aspect of system design, new challenges arise in this area of research.

Beside the technical contributions, this special issue also offers a project description by Schmid and Finzel that exploits mutual explanations for interactive learning as part of an interdisciplinary research project on transparent machine learning in the area of decision support in healthcare. The project combines deep learning black-box approaches with interpretable machine learning for classification of medical images and allows for high predictive accuracy enabled by deep learning but offers transparency and comprehensibility of interpretable models at the same time.

Moreover, Sokol and Flach discuss the promise of interactive explanations for machine learning transparency. This discussion is highly recommendable to both researchers and practitioners who are interested in this area because it can provide very valuable insights.

3 Outlook

It is becoming progressively clearer that interactive machine learning will play a central role in the upcoming Artificial Intelligence revolution. There is a very simple reason for this: in order to be beneficial in the real world, AI agents will have to be able to communicate and collaborate with, learn from, and teach to its inhabitants: us. Finding the right ways of injecting learning in human-machine interaction is a prerequisite to opening up many scientific and commercial opportunities in AI, from recommender systems that fully grasp the customer's intentions, to personal aids that think similarly to their users, to machines that educate children based on their perceived competence level and engagement.

Clearly, the topic of algorithmic transparency, explainability and fairness becomes even more important when combined with interactive machine learning. Different forms of visualisations and degrees of transparency features (e.g., [9]) may exert different influences on the human counterpart in interactive machine learning. Subsequently, this may greatly affect internalization of knowledge on the side of the human user, or actions taken by the human to teach the machine. For example, if a machine only provides metrics such as accuracy, this could lead an inexperienced user to make the strong assumption that the model also has a high recall and precision score. Furthermore, when systems present actual feature importance to the users, these may develop different notions towards the system regarding fairness, traceability and overall logic of the inferences. Thus, the topic of transparency and its effects on interactive machine learning processes must be studied in more detail, as it may be a decisive factor leading to either beneficial or detrimental effects in interactive machine learning.

We thank all authors and reviewers for their contributions, which made this special issue possible, and hope that this issue will spur further interest in interactive machine learning.

4 Content

In summary, this special issue compiles the following content.

4.1 Surveys

- *Power to the Oracle? Design Principles for Interactive Labeling Systems in Machine Learning*
  Mario Nadj, Merlin Knaeble, Maximilian Xiling Li and Alexander Maedche

4.2 Technical Contributions

- *eXplainable Cooperative Machine Learning with NOVA*
  Tobias Baur, Alexander Heimerl, Florian Lingenfelser, Johannes Wagner, Michel F. Valstar, Bjoern Schuller and Elisabeth André
- *Active and Incremental Learning with Weak Supervision*
  Clemens-Alexander Brust, Christoph Käding, Joachim Denzler
- *Interactive Transfer Learning in Relational Domains*
  Raksha Kumaraswamy, Nandini Ramanan, Phillip Odom, Sriraam Natarajan
- *Measuring the Quality of Explanations: The System Causability Scale (SCS)*
  Andreas Holzinger, André Carrington, Heimo Müller
- *How and what can Humans Learn from being in the Loop?*
  Benjamin M. Abdel-Karim, Nicolas Pfeuffer, Gernot Rohde, Oliver Hinz

4.3 Project Reports

- *Mutual Explanations for Cooperative Decision Making in Medicine*
  Ute Schmid and Bettina Finzel

4.4 Discussion

- *One Explanation Does Not Fit All – The Promise of Interactive Explanations for Machine Learning Transparency*
  Kacper Sokol and Peter Flach

4.5 Dissertations

- *Dealing with Mislabeling via Interactive Machine Learning*
  Wanyi Zhang, Andrea Passerini and Fausto Giunchiglia

5 Service

Finally, let us provide some links to additional material such as relevant conferences and journals as well as further examples for research groups on interactive machine learning.

5.1 Conferences and Workshops

- International Joint Conference on Artificial Intelligence, https://ijcai20.org/
- AAAI Conference on Artificial Intelligence, https://aaai.org/Conferences/AAAI-21/
- ACM CHI, https://chi2021.acm.org/
- ACM Conference on Fairness, Accountability, and Transparency, https://facctconference.org/
- AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, https://www.aies-conference.com/
- European Conference on Artificial Intelligence, http://ecai2020.eu
- European Conference on Information Systems, https://ecis2020.ma/

5.2 Journals

- Artificial Intelligence, https://www.journals.elsevier.com/artificial-intelligence
- Journal of Artificial Intelligence Research, https://jair.org/index.php/jair
- Machine Learning, https://www.springer.com/journal/10994
- Journal of Machine Learning Research, http://www.jmlr.org
- Machine Learning and Knowledge Extraction, https://www.mdpi.com/journal/make
- Business & Information Systems Engineering, https://www.springer.com/journal/12599

5.3 Working Groups

- Human-Centered AI Group, Medical University Graz, https://human-centered.ai/
- People + AI (PAIR) Group, Google Research, https://www.aclweb.org/anthology/P19-1234/
- Interactive Machine Learning Lab, DFKI - German Research Center for Artificial Intelligence, http://iml.dfki.de
- Machine Teaching Group, Microsoft Research, https://www.microsoft.com/en-us/research/group/machine-teaching-group/s

– Artificial Intelligence and Machine Learning Lab, TU Darmstadt, https://www.ml.informatik.tu-darmstadt.de

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Amershi S, Cakmak M, Knox WB, Kulesza T (2014) Power to the people: the role of humans in interactive machine learning. Ai Mag 35(4):105–120
2. Angluin D (1988) Queries and concept learning. Mach Learn 2(4):319–342
3. Boult T, Cruz S, Dhamija A, Gunther M, Henrydoss J, Scheirer W (2019) Learning and the unknown: surveying steps toward open world recognition. Proc AAAI Conf Artif Intell 33:9801–9807
4. Boutilier C (2002) A POMDP formulation of preference elicitation problems. In: AAAI/IAAI, pp 239–246
5. Dawid AP, Skene AM (1979) Maximum likelihood estimation of observer error-rates using the em algorithm. J R Stat Soc Ser C (Appl Stat) 28(1):20–28
6. Dietterich TG (2017) Steps toward robust artificial intelligence. AI Mag 38(3):3–24
7. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2018) A survey of methods for explaining black box models. ACM Comput Surv (CSUR) 51(5):1–42
8. Hinton G, LeCun Y, Bengio Y (2015) Deep learning. Nature 521(7553):436–444
9. Ribeiro M.T, Singh S, Guestrin C (2016) "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1135–1144
10. Russell S (2019) Human compatible: artificial intelligence and the problem of control. Penguin, London
11. Schölkopf B (2019) Causality for machine learning. arXiv preprint arXiv:1911.10500
12. Scholz M, Dorner V, Franz M, Hinz O (2015) Measuring consumers' willingness to pay with utility-based recommendation systems. Decis Support Syst 72:60–71
13. Scholz M, Franz M, Hinz O (2017) Effects of decision space information on MAUT-based systems that support purchase decision processes. Decis Support Syst 97:43–57
14. Settles B (2012) Active learning. Morgan & Claypool, San Rafael
15. Smith B, Linden G (2017) Two decades of recommender systems at amazon.com. IEEE Internet Comput 21(3):12–18
16. Teso S, Kersting K (2019) Explanatory interactive machine learning. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp 239–245