Interactive Machine Learning (iML): a challenge for Game-based approaches

Andreas Holzinger, Markus Plass, Michael D. Kickmeier-Rust
Holzinger Group HCI-KDD, Institute for Medical Informatics, Statistics and Documentation
Medical University Graz, Austria
Knowledge Technologies Institute, Graz University of Technology, Austria
a.holzinger@hci.kdd.org

2-page Extended Abstract CiML-2016-Workshop, Barcelona, Spain

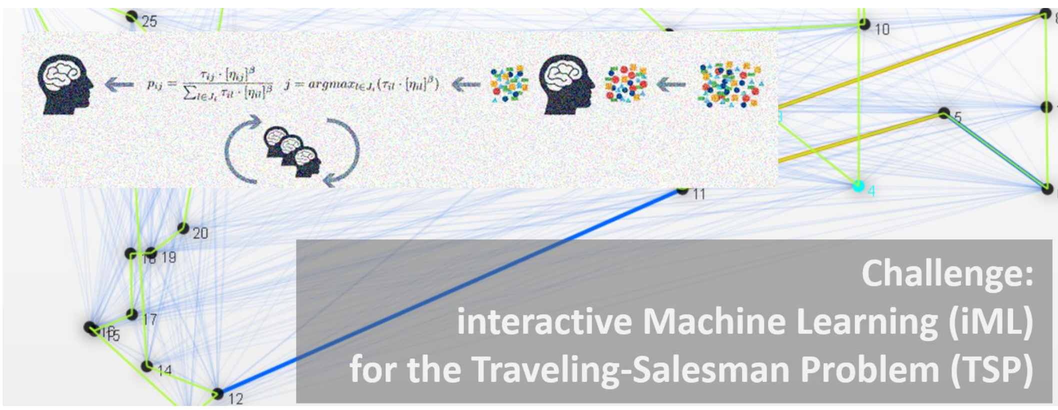

Figure 1: iML can be defined as algorithms that interact with agents and can optimize their learning behavior through these interactions, where the agents can also be humans [1]. These humans-in-the-loop can be beneficial in solving computationally hard problems

The goal of the ML-community is to design and develop algorithms which can learn from data and improve with experience over time. However, the application of such automatic machine learning (aML) approaches in the complex biomedical domain seems elusive in the near future, and a good example are Gaussian processes, where aML (e.g. standard kernel machines) struggle on function extrapolation problems - which are trivial for human learners [2]. Psychological research indicates that *human intuition* is not an 'esotheric' concept; much more, it is based on distinct behavioral and cognitive strategies that developed evolutionary over millions of years [3]. For improving ML, we need to identify the concrete mechanisms and we argue that this can be done best *by observing crowd behaviors and decisions in gamified situations*.

We have proved experimentally that the iML approach can be used to advance current Traveling Salesman Problem (TSP) solving methods [4]. TSP is important, as it appears in a number of practical problems in biomedical informatics, e.g. the native folded three-dimensional conformation of a protein is its lowest free energy state and both a two- and three-dimensional folding processes as a free energy minimization problem belong to a large set of computational problems, assumed to be very hard (conditionally intractable) [5]. For our experiment we applied the Ant Colony Optimization (ACO) - which is a fully automatic (aML) approach without any human intervention: The ants walk around and update the global weights after each iteration. This procedure is repeated a distinct number of times. Now, following the iML-approach the human can open the black-box and manipulate the algorithm by changing the behavior of the ants; based on the Inner Ant System for TSP [6] this could be understood as adding or removing of pheromones on a trail between nodes of a graph. Consequently, the trails become more or less interesting for the ants and there is a high probability that they will consider trail changes. Due to the fact that we have previous experience with gamification for education [7] we are currently gamifying our experiment: To achieve an intuitive interaction the problem is presented as a game, where the user has to reduce the length of trail, by changing the amount of pheromones on the lines between the nodes. The implementation is based on Java-Script, with the advantage of being a browser based solution with the benefit of platform independence and no installation requirements. For the test-setup we used 30 ants, 250 iterations. For the other parameters including evaporation we choose fixed default values, this makes it easier to compare the results. Humans performance is excellent for problems with a size of about 100 nodes these could be solved by humans intuitively, for problems with x-thousands of nodes it is impossible to get an overview and to find useful solutions within acceptable time. The idea is to split a large 2D-problem into smaller overlapping sub-problems which can be solved separately, and through the overlap they can be merged after the interaction together, this approach will allow us to work also with large data sets from the biomedical domain (e.g. [8]).

The challenge is in 2D-visualization, i.e. to represent the data, particularly larger and complex data sets, in a form which is still comprehensible by a human. For this purpose we are administering a software engineering student challenge within our university course at our faculty of computer science; but we could also extent this challenge to international level. By representing the data in form of easy to understand 2D graphics, humans may find possible solutions quickly and intuitively. By this means, distinct decision and problem solving patterns can be identified, which so far have been considered being this vague concept of human intuition, hence can contribute to the outcomes of the aML-Algorithm in an interactive way. By this approach we believe to help to answer the difficult question of ”What is interesting?” to find a better solution than the aML-approach would find in the same time, given that this time span is reasonably limited. Our first analyses showed, that on two of the three test data sets the iML approach gives better results than the aML approach, this results are reached by only 2.4 (in mean) human interactions.

References

[1] A Holzinger. Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics*, 3(2):119–131, 2016. doi: 10.1007/s40708-016-0042-6.

[2] A G Wilson, C Dann, C Lucas, and E P Xing. The human kernel. In Corinna Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems, NIPS 2015*, volume 28, pages 2836–2844, 2015.

[3] M D Lieberman. Intuition: a social cognitive neuroscience approach. *Psychological Bulletin*, 126(1):109–137, 2000. doi: 10.1037/0033-2909.126.1.109.

[4] A Holzinger, M Plass, K Holzinger, GC Crisan, CM Pintea, and V Palade. Towards interactive machine learning (iml): Applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In *Springer Lecture Notes in Computer Science LNCS 9817*, pages 81–95. Springer, Heidelberg, Berlin, New York, 2016. doi: 10.1007/978-3-319-45507-56.

[5] P Crescenzi, D Goldman, C Papadimitriou, A Piccolboni, and M Yannakakis. On the complexity of protein folding. *Journal of computational biology*, 5(3):423–465, 1998. doi: 10.1016/S0092-8240(05)80170-3.

[6] Camelia Pintea, D Dumitrescu, and PC Pop. Combining heuristics and modifying local information to guide ant-based search. *Carpathian Journal of Mathematics*, 24(1):94–103, 2008.

[7] M Ebner and A Holzinger. Successful implementation of user-centered game based learning in higher education: An example from civil engineering. *Computers and Education*, 49(3): 873–890, 2007. doi: 10.1016/j.compedu.2005.11.026.

[8] I Guyon, D Battaglia, A Guyon, V Lemaire, J G Orlandi, B Ray, M Saeed, J Soriano, A Statnikov, and O Stetter. Design of the first neuronal connectomics challenge: From imaging to connectivity. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 2600–2607. IEEE, 2014.