

Open Machine Learning 2017

9 – 13 October 2017 @ Lorentz Center

The rise of Machine Learning and Artificial Intelligence in nearly all aspects of society stands in stark contrast with the irreproducibility of the literature and its inaccessibility to the general public. Therefore, we have developed OpenML, an online collaborative science platform for Machine Learning, available at <http://openml.org>. The aim of this workshop is to bring together developers and users of the platform to stimulate further development.



During the workshops and the weeks after that, we:

- Started the Open Machine Learning Foundation to support the development of the platform (organizationally and financially). It will allow collaborations with companies;
- Started collaboration with Microsoft to offer free scientific notebook hosting to all OpenML users, and started talks on supporting large-scale machine learning benchmarks;
- We established new benchmarks for machine learning algorithms, and in the process we cleaned up hundreds of datasets. A paper will be submitted to JMLR soon;
- Drastically improved Python support + integration into the popular scikit-learn library;
- Grew a stronger community of young enthusiastic scientists that continue to collaborate;
- Developed training materials + guides to introduce many more people to the platform. Heidi Seibold also presented OpenML during the 'this week's discoveries' lectures;
- Created blueprints for dataset quality checking, automated data preprocessing, deep learning, model interpretability, clustering, RAMPs,... that we will follow up on over the next months.

The philosophy of OpenML is frictionless online collaboration, and thus accelerate machine learning and data driven discovery. One key aspect of this is standardized benchmarking, to make sure that there is actual scientific progress (not just hype or 'benchmarking'). The workshop showed that, both from academic and corporate perspective, a more directed effort to streamline algorithm performance evaluation is needed, and we've made substantial progress into supporting this through the OpenML platform. We plan to publish this soon and are working on community awareness and adoption.

The workshop was organized as a *hackathon*, an event where participants present their goals and ideas, and then work on them in small teams for many hours or days at a time. We especially focussed on user engagement, through high-profile speakers (including Microsoft, Amazon, H2O, the Turing Institute, Paris-Saclay) and clear tutorials. Most new people were very active in the breakouts and broadened our perspective. Some people found that the breakouts were too interesting (and too many), in that they spent less time on actual coding.

Here are a few abstracts from the overwhelmingly positive feedback:

- Thanks for organizing the workshop! It turned out to be quite fruitful, I made valuable new connections;
- I learned new methods and advancements of machine learning; applications in other fields than my field;
- The benchmarking discussions were really useful;
- I loved the format. However, the breakouts were so popular that that breakout room, 301, was too small;
- Talks were good, and not too much. But there was too much discussion (breakouts) and too little coding;
- Format is fine but comes to its limit as the attendance grows. If it keeps getting bigger, consider splitting into "user" and "developer" workshops.

Altogether, we think the workshop was a great success and it pushed the idea of Open Science in Machine Learning forward by the continued development of OpenML.

Bernd Bischl (Munich, Germany)

Jan van Rijn (Leiden, Netherlands)

Joaquin Vanschoren (Eindhoven, The Netherlands)