AutoML 2015 workshop @ ICML 2015

Rapport sur les contributions

ID de Contribution: **0**   Type: **Non spécifié**

Invited Talk: Algorithm Recommendation as Collaborative Filtering

samedi 11 juillet 2015 10:30 (40 minutes)

Orateur: SEBAG, Michele (CNRS)

Classification de Session: Session 2

ID de Contribution: **1**   Type: **Non spécifié**

Poster spotlights 2

samedi 11 juillet 2015 11:10 (18 minutes)

9 spotlights of 2 minutes each

Classification de Session: Session 2

ID de Contribution: **2** Type: **Non spécifié**

1st Poster Session

samedi 11 juillet 2015 11:30 (30 minutes)

Classification de Session: Session 2

Real time data analysis at the LHC : present and future

The large hadron collider (LHC), which collides protons at an energy of 14 TeV (for non-physicists, each beam of protons carries roughly the energy of a TGV train going at full speed), produces hundreds of exabytes of data per year, making it one of the largest sources of data in the world today. At present it is not possible to even transfer most of this data from the four main particle detectors at the LHC to "offline" data facilities, much less to permanently store it for future processing. For this reason the LHC detectors are equipped with real-time analysis systems, called triggers, which process this volume of data and select the most interesting proton-proton collisions. The LHC experiment triggers reduce the data produced by the LHC by between 1/1000 and 1/10000, to tens of petabytes per year, allowing its economical storage and further analysis. The bulk of this data-reduction is performed by custom electronics which ignores most of the data in its decision making, and is therefore unable to exploit the most powerful known data analysis strategies developed by e.g. the machine learning community. In this talk I will cover the present status of real-time data analysis at the LHC, before explaining why the future upgrades of the LHC experiments will increase the volume of data which can be sent off the detector and into off-the-shelf data processing facilities (such as CPU or GPU farms) to tens of exabytes per year. This development will simultaneously enable a vast expansion of the physics programme of the LHC's detectors, and make it mandatory to develop and implement a new generation of real-time multivariate analysis tools in order to fully exploit this new potential of the LHC. I will explain what work is ongoing in this direction and hopefully motivate why more effort is needed in the coming years.

ID de Contribution: **4**   Type: **Non spécifié**

Invited Talk: Open Research Problems in AutoML

samedi 11 juillet 2015 08:30 (40 minutes)

Orateur: CARUANA, Rich (Microsoft Research)

Classification de Session: Session 1

ID de Contribution: **5**   Type: **Non spécifié**

Invited Talk: Bandits and Bayesian optimization for AutoML

samedi 11 juillet 2015 09:10 (40 minutes)

Complex optimization and decision making tasks are beginning to play an increasingly crucial role across a wide variety of scientific fields. This is becoming more and more evident as entire research programs are being automated.

In this talk I'll describe a set of methods, known as Bayesian optimization, which provide a very sample efficient approach to this problem. Much of the gains of these methods are obtained by building a posterior model of a function during optimization in order to efficiently explore its surface. I will further describe a number of advanced search mechanisms and models and show how these can be used for automating Machine Learning problems. Finally, I will also briefly provide links to related bandit literature.

Orateur: HOFFMANN, Matthew (University of Cambridge)

Classification de Session: Session 1

ID de Contribution: **6**   Type: **Non spécifié**

Poster Spotlights 1

samedi 11 juillet 2015 09:50 (10 minutes)

5 spotlights of 2 minutes each

Classification de Session: Session 1

ID de Contribution: **7**   Type: **Non spécifié**

Invited Talk: OpenML: A Foundation for Networked & Automatic Machine Learning

samedi 11 juillet 2015 16:30 (40 minutes)

OpenML is an online machine learning platform where scientists can automatically log and share data sets, code, and experiments, organize them online, and collaborate with researchers all over the world. It helps to automate many tedious aspects of research, is readily integrated into several machine learning tools, and offers easy-to-use APIs. It also enables large-scale and real-time collaboration, allowing researchers to build directly on each other's latest results, and track the wider impact of their work. Ultimately, this provides a wealth of information for building systems that learn from previous experiments, to either assist people while analyzing data, or automate the process altogether.

Orateur: VANSCHOREN, Joaquin (Eindhoven University of Technology)

Classification de Session: Session 4

ID de Contribution: **8**   Type: **Non spécifié**

AutoML Challenge

samedi 11 juillet 2015 17:10 (20 minutes)

Orateur: BOULLE, Marc (Orange)

Classification de Session: Session 4

ID de Contribution: **9**

Type: **Non spécifié**

Panel Discussion: Next steps for AutoML

samedi 11 juillet 2015 17:30 (30 minutes)

Panelists: Marc Boulle, Rich Caruana, David Duvenaud, Matthew Hoffmann, Juergen Schmidhuber, Michèle Sebag, Joaquin Vanschoren.

Classification de Session: Session 4

ID de Contribution: **10**   Type: **Non spécifié**

Invited Talk: Recursive Self-Improvement

samedi 11 juillet 2015 14:00 (40 minutes)

Most machine learning researchers focus on domain-specific learning algorithms. Can we also construct meta-learning algorithms that can learn better learning algorithms, and better ways of learning better learning algorithms, and so on, restricted only by the fundamental limitations of computability? In 1965, J. Good already made informal remarks on an intelligence explosion through such recursive self-improvement (RSI).

I will discuss various concrete algorithms (not just vague ideas) for RSI: 1. My diploma thesis (1987) proposed an evolutionary system that learns to inspect and improve its own learning algorithm, where Genetic Programming (GP) is recursively applied to itself, to invent better learning methods, meta-learning methods, meta-meta-learning etc. 2. RSI based on the self-referential Success-Story Algorithm for self-modifying probabilistic programs (1997) was already able to solve complex tasks. 3. My self-referential deep recurrent neural networks (since 1993) run and inspect and change their own weight change algorithms. Back in 2001, my former student Hochreiter (now prof) already had a practical implementation of such an RNN that meta-learns an excellent learning algorithm, at least for a limited domain. 4. The Goedel machine (2006) is the first RSI that is mathematically optimal in a particular sense. Will RSI finally take off in the near future?

Orateur: SCHMIDHUBER, Juergen (IDSIA)

Classification de Session: Session 3

ID de Contribution: **11**
Type: **Non spécifié**

Invited Talk: Automatically constructing models, and automatically explaining them, too.

samedi 11 juillet 2015 14:40 (40 minutes)

How could an artificial intelligence do statistics? It would need an open-ended language of models, and a way to search through and compare those models. Even better would be a system that could explain the different types of structure found, even if that type of structure had never been seen before. This talk presents a prototype of such a system, which builds structured Gaussian processes regression models by combining covariance kernels to build a custom model for each dataset. The resulting models can be broken down into relatively simple components, and surprisingly, it's not hard to write code that automatically describes each component, even for novel combinations of kernels. The result is a procedure that takes in a dataset, and outputs a report with plots and English descriptions of the different types of structure found in that dataset.

Orateur: DUVENAUD, David (Harvard University)

Classification de Session: Session 3

ID de Contribution: **12** Type: **Non spécifié**

2nd Poster Session

samedi 11 juillet 2015 15:20 (40 minutes)

Classification de Session: Session 3