Bayesian Optimization for More Automatic Machine Learning

(extended abstract for invited talk)

Frank Hutter[1]

Bayesian optimization (see, e.g., [2]) is a framework for the optimization of expensive blackbox functions that combines prior assumptions about the shape of a function with evidence gathered by evaluating the function at various points. In this talk, I will briefly describe the basics of Bayesian optimization and how to scale it up to handle structured high-dimensional optimization problems in the sequential model-based algorithm configuration framework SMAC [6].

Then, I will discuss applications of SMAC to two structured high-dimensional optimization problems from the growing field of *automatic machine learning*:

- Feature selection, learning algorithm selection, and optimization of its hyperparameters are crucial for achieving good performance in practical applications of machine learning. We demonstrate that a combined optimization over all of these choices can be carried out effectively by formulating the problem of finding a good instantiation of the popular WEKA framework as a 768-dimensional optimization problem. The resulting Auto-WEKA framework [7] allows non-experts with some available compute time to achieve state-of-the-art learning performance on the push of a button.
- Deep learning has celebrated many recent successes, but its performance is known to be very sensitive to architectural choices and hyperparameter settings. Therefore, so far its potential could only be unleashed by deep learning experts. We formulated the combined problem of selecting the right neural network architecture and its associated hyperparameters as a 81-dimensional optimization problem and showed that an automated procedure could find a network whose performance exceeded the previous state-of-the-art achieved by human domain experts using the same building blocks [3]. Computational time remains a challenge, but this result is a step towards deep learning for non-experts.

To stimulate discussion, I will finish by highlighting several further opportunities for combining meta-learning and Bayesian optimization:

- Prediction of learning curves [3],
- Learning the importance of hyperparameters (and of meta-features) [4, 5], and
- Using meta-features to generalize hyperparameter performance across datasets [1, 8], providing a prior for Bayesian optimization.

Based on joint work with Tobias Domhan, Holger Hoos, Kevin Leyton-Brown, Jost Tobias Springenberg, and Chris Thornton.

[1] University of Freiburg, Germany. Email: fh@cs.uni-freiburg.de.

REFERENCES

[1] R. Bardenet, M. Brendel, B. Kgl, and M. Sebag, 'Collaborative hyperparameter tuning', in *Proc. of ICML*, (2013).

[2] E. Brochu, V. M. Cora, and N. de Freitas, 'A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning', *CoRR*, **abs/1012.2599**, (2010).

[3] Tobias Domhan, Tobias Springenberg, and Frank Hutter, 'Extrapolating learning curves of deep neural networks', in *ICML 2014 AutoML Workshop*, (June 2014).

[4] F. Hutter, H. Hoos, and K. Leyton-Brown, 'Identifying key algorithm parameters and instance features using forward selection', in *Learning and Intelligent Optimization*, pp. 364–381, (2013).

[5] F. Hutter, H. Hoos, and K. Leyton-Brown, 'An efficient approach for assessing hyperparameter importance', in *International Conference on Machine Learning*, (2014).

[6] F. Hutter, H. H. Hoos, and K. Leyton-Brown, 'Sequential model-based optimization for general algorithm configuration', in *Proc. of LION-5*, (2011).

[7] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, 'Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms', in *Proc. of KDD'13*, (2013).

[8] D. Yogatama and G. Mann, 'Efficient transfer learning method for automatic hyperparameter tuning', in *Proc. of AISTATS*, (2014).