

Preface: On-line Trading of Exploration and Exploitation 2

Dorota Głowacka, Louis Dorard, John Shawe-Taylor

The second edition of the *On-line Trading of Exploration and Exploitation Workshop* took place on 2 July 2011 in Bellevue, Washington and was co-located with the *28th International Conference on Machine Learning*. The proceedings contain a selection of papers presented at the workshop.

On-line problems such as website optimisation require to trade exploration and exploitation in order to learn and optimise an unknown target. For instance, in the *Pascal Exploration & Exploitation Challenge 2011*, a web server observes clicks from visitors to a webpage, and aims to maximise the ratio of clicks per page views. The relationship between the visitor-content pairs and clicks is unknown but it can be learnt from past observations. The content presented to the visitors is either chosen to improve our model of clicks in regions of uncertainty or it is based on the model that has been built so far. These two distinct motives are referred to as exploration and exploitation. Thus, the web server should explore enough to be able to build an accurate model of the visitor's behaviour, while allowing for sufficient exploitation in order to earn clicks.

The problem of trading exploration and exploitation was first tackled in the "multi-armed bandit" formalism, in which the target observations are compared to rewards obtained when pulling arms of slot-machines. The first theoretical analysis focused on independent reward distributions for each arm. However, in a real-world scenario, arms (inputs) are rarely independent and modelling the dependencies is essential in order to obtain the best learning rate. Gaussian Processes (GP), for instance, are a powerful modelling tool that has been widely used in bayesian online optimisation, in combination with heuristics such as the Most Probable Improvement for selecting inputs where to sample the function. However, performance guarantees for the use of GP in a bandit setting were only found in 2010, when used in combination with the Upper Confidence Bound heuristic for trading exploration and exploitation.

The exploration/exploitation dilemma is a recurrent topic in many areas of research, e.g. global optimisation, reinforcement learning, tree search, recommender systems or information retrieval. The trading of exploration and exploitation is particularly of high importance in various large-scale applications, such as sponsored search advertising or content-based information retrieval, where the aim is to help users quickly access the information they are looking for. The workshop provided an opportunity to present, compare and discuss the performance of different exploration/exploitation techniques as well as theoretical analysis of such algorithms. A particular focus of the workshop were large scale applications. The results of the *Pascal Exploration & Exploitation Challenge 2011* were also presented during the workshop.