Active Automatic Algorithm Selection Using Multi-armed Bandits

Hans Degroote and Patrick De Causmaecker
KU Leuven Department of Computer Science
CODeS & iMinds-ITEC

What is automatic algorithm selection?

* Algorithm selection is modelled as a classification problem:
  - Divide the instance space into subsets of similar instances (based on the features)
  - Learn for each subset the best algorithm

Current offline solution approach

* Training data is fed to a supervised learning technique to create the selection mapping
* Training data is the performance of all algorithms on a set of training instances

Situating this research

* A continuous stream of performance data is generated when doing algorithm selection
* This free data is not used to improve the selection mapping by current algorithm selection approaches

Is it possible to transfer knowledge between bandits?

* Biased samples are produced by
  - Offline data
  - Online data based on which the subset structure has been modified
* The performance data points by which a new split is inspired are not random samples for the two newly defined subsets.
* The proposed curtailment-inspired approach for modifying the subset structure ignores biased samples: the data used to define a new subset is not used to help learn the best algorithm for the new subset.
* However, the very reason the subset was created was because one algorithm appeared better. This algorithm should be given preference.

Questions to the multi-armed bandit community:

* Is it possible to incorporate the information contained in biased samples into a multi-armed bandit problem?
* How can the expected reward of a bandit process be calculated?
* What is the link between Bayesian inference/updating and multi-armed bandits?

Conclusions

* An active learning methodology for automatic algorithm selection has been outlined
* At its core it relies on solving multiple multi-armed bandit problems
* An extension splits the reward distribution of a bandit, thereby introducing two new bandits for which biased samples are available that cannot be directly used, yet contain relevant information
* An additional complication is that the reward distributions are often unknown and might even have an ill-defined mean

Algorithm selection with multi-armed bandits:

* After the training phase the instance space is divided into subsets and a best algorithm has been learned for each subset, but it might not be the actual best.
* Based on the feedback generated during the online phase, each subset faces its own active learning problem: “How can the best algorithm for my instance distribution be learned?”?
* More precisely, “What is the regret-optimal policy for selecting an algorithm to solve sequential random samples from my instance distribution?”?
* This is a multi-armed bandit problem at heart:
  - arms = algorithms
  - pulling an arm = selecting an algorithm
  - reward = algorithm performance

Active learning for automatic algorithm selection

* Identify the subset to which an instance belongs
* Select the algorithm that is best according to a solution to the subset’s multi-armed bandit problem
* Feed the performance of the selected algorithm back to the subset’s bandit

Challenges

* The type of distribution underlying algorithm performance is often unknown
* The distributions underlying algorithm performance might not have a well-defined mean (heavy-tailed distributions)

Contact

Hans Degroote
KU Leuven
Email: Hans.Degroote@kuleuven-kulak.be
Phone: +32 56 24 62 88