

AlphaGo goes shopping

Peggy Cellier, IRISA - SemLIS, INSA Rennes, peggy.cellier@irisa.fr

Romarc Gaudel, CREST, ENSAI, romarc.gaudel@ensai.fr

Alexandre Termier, IRISA - Lacodam, Univ. de Rennes 1 Alexandre.Termier@irisa.fr

Location: IRISA, team SemLIS, Rennes

Keywords: data science, data mining, pattern mining, Minimum Description Length (MDL), machine learning, Monte Carlo Tree Search (MCTS)

Context and description of the internship:

Data mining is the process that allows identifying knowledge from data [1]. Among the data mining techniques, pattern mining, as association rule mining or sequential pattern mining, is the one that extracts regularities from data. That kind of techniques has been used in several fields as for instance marketing to find products sold together or linguistics to study the stylistics of authors. A well-known drawback of pattern mining is the risk of overwhelming the analyst with millions of patterns. Several approaches have been proposed to address that problem, the most promising are based on the minimum description length (MDL) principle [2] (ex: KRIMP [3, 4]). In those approaches, a subset of the extracted patterns is selected thanks to information theoretic criteria. The search space is huge, however its exploration is currently handled by relatively straightforward greedy algorithms.

On the other hand, Monte Carlo Tree Search (MCTS) [5] is a search algorithm for decision processes exhibiting huge search spaces. It is looking for the best leaf in a tree (and the path to that leaf). Roughly, it gradually expands a search tree which corresponds to promising branches. This approach was at the root of first master-level artificial Go-players in 2008 [6] and was one of the main ingredients of AlphaGo [7] (who did beat the Go world champion in 2016).

The aim of the internship is to explore how the pattern selection step of KRIMP can be improved by exploiting the capability of MCTS to explore a wide search space. The expected results is to find higher quality subsets of patterns (according to MDL criteria). The approach will be validated experimentally on synthetic and real datasets from our applicative partners (retail industry, cybersecurity, agronomy, bioinformatics among others).

This internship requires to have at the same time a taste for theory (MDL), algorithms (MCTS) and experiments.

[1] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth. **From Data Mining to Knowledge Discovery in Databases**. AI Magazine, 1996.

[2] P. Grünwald. **Model Selection Based on Minimum Description Length**. Journal of Mathematical Psychology, 2000.

[3] A. Siebes, J. Vreeken, M. van Leeuwen. **Item Sets that Compress**. Proceedings of SDM'06.

[4] M. van Leeuwen, J. Vreeken, A. Siebes. **Compression Picks the Item Sets that Matter**. ECML PKDD'06.

[5] L. Kocsis; C. Szepesvári. **Bandit based Monte-Carlo Planning**. Proceedings of ECML'06.

- [6] S. Gelly, D. Silver. **Achieving Master Level Play in 9 x 9 Computer Go**. Proceedings of AAAI'08, 1537-1540
- [7] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis. **Mastering the game of Go with deep neural networks and tree search**. Nature 529, 484–489 (28 January 2016).