Semantic Graph Mining for Black-Box Optimisation

Master internship

March 2023 - September 2023

Supervisors
Carola Doerr, LIP6, RO team
Marie-Jeanne Lesot, LIP6, LFI team
Claire Laudy, Thales TRT
Workplace: LIP6, Sorbonne University, 4 place Jussieu, Paris
To apply: Send a CV, a short motivation letter and a transcript/summary of obtained grades to the supervisors

The general aim of the internship is to exploit expert knowledge regarding properties of optimisation algorithms and problems, represented in the formal frameworks of ontologies and conceptual graphs, and to **develop tools to automatically extract underlying correlations:** the objective is to allow understanding the reasons why an algorithm is more appropriate than others to solve a problem depending on its characterisation and possibly to offer new tools to configure optimisation algorithms.

In this context, the internship work will address the task of **frequent pattern mining** for data represented as conceptual graphs, and apply such algorithms to real data representing knowledge about optimisation algorithms and problems.

Regarding the considered application, the work aims at processing OPTION, the *OPTImisation algorithm benchmarking ONtology*[1] (Kostovska et al. 2021; 2022): OPTION groups and annotates benchmarking data from black-box optimisation, hosting more than 200 algorithm performance data from the BBOB collection of the COCO framework (Hansen et al., 2020) and from various benchmark suits of the Nevergrad environment (Rapin & Teytaud, 2018). OPTION provides the vocabulary needed for semantic annotation of the core entities involved in the process of benchmarking black-box optimisation algorithms, such as algorithms, problems and evaluation measures. These very rich data can be represented in the formalism of conceptual graphs (see e.g. Chein & Mugnier, 2008) which offers the advantage of combining a logical semantic and efficient manipulation tools based on graph theory.

Extracting knowledge from conceptual graphs can take the form of identifying frequent patterns, i.e. subgraphs that occur frequently and can thus be interpreted as relevant regularities. This relates to the domain of Frequent Subgraph Mining, for which numerous approaches have been proposed (see e.g. Jiang et al., 2013). Dedicated algorithms have been proposed for the specific case of taxonomy labelled graphs (Inokuchi et al., 2000; Cakmak & Ozsoyoglu, 2008; Petermann et al., 2017), that exploit the particular characteristics of these types of graphs to improve both their efficiency and the relevance of the extracted patterns. Recent works conducted in the LFI team focused further on the case of conceptual graphs: the cgSpan algorithm (Faci et al., 2021) exploits the information about relation arity to decrease efficiently the number of explored pattern candidates, the relation signatures to prune redundant pattern candidates with little informativeness as well as inference rules to extend candidates faster. The goal of the internship is to build upon that work and to extend it further.

The internship work aims at designing and implementing a comparison tool of algorithms of frequent pattern mining and proposing new approaches for the case of conceptual graphs. After validating the propositions on baseline and synthetic data, the second part of the internship will turn to the real case of understanding black-box optimisation applying the approach to the OPTION data and developing an XAI tool for this case.

[1] available at the BioPortal:
https://bioportal.bioontology.org/ontologies/OPTION-ONTOLOGY

Candidate profile: Background in a quantitative field such as Computer Science, Engineering, Statistics, Operations Research, Mathematics is required. We expect willingness to conduct empirical research as well as experience with the python programming language. Knowledge of French is not required, our working languages are French and English. International students are hence very welcome to apply.

References

Cakmak, A., & Ozsoyoglu, G. (2008). Taxonomy-superimposed graph mining. *Proc. of the 11th Int. Conf. on Extending database technology: Advances in database technology* (pp. 217–228).

Chein, M., & Mugnier, M.-L. (2008). *Graph-based knowledge representation: Computational foundations of conceptual graphs*. Springer.

Faci, A., Lesot, M.-J., & Laudy, C. (2021). cgspan: Pattern mining in conceptual graphs. *Proc. of the Int. Conf. on Artificial Intelligence and Soft Computing* (pp. 149–158). Springer, LNCS12855.

Hansen, N., Auger, A., Ros, R., Mersmann, O., Tušar, T., & Brockhoff, D. (2020). COCO: A platform for comparing continuous optimizers in a black-box setting. *Optimization Methods and Software*, *36*, 114–144.

Inokuchi, A., Washio, T., & Motoda, H. (2000). An apriori-based algorithm for mining frequent substructures from graph data. *Proc. of the European Conf. on principles of data mining and knowledge discovery* (pp. 13–23).

Jiang, C., Coenen, F., & Zito, M. (2013). A survey of frequent subgraph mining algorithms. *The Knowledge Engineering Review*, *28*, 75–105.

Kostovska, A., Vermetten, D., Doerr, C., Dzeroski, S., Panov, P., & Eftimov, T. (2021). Option: optimization algorithm benchmarking ontology. *Proc. of the Genetic and Evolutionary Computation Conference Companion, GECCO Companion 2021* (pp. 239–240).

Kostovska, A., Vermetten, D., Doerr, C., Dzeroski, S., Panov, P., & Eftimov, T. (2022). OPTION: optimization algorithm benchmarking ontology. *CoRR*, *abs/2211.11332*.

Petermann, A., Micale, G., Bergami, G., Pulvirenti, A., & Rahm, E. (2017). Mining and ranking of generalized multi-dimensional frequent subgraphs. *Proc. of the 12th Int. Conf. on Digital Information Management (ICDIM)* (pp. 236–245).

Rapin, J., & Teytaud, O. (2018). Nevergrad - A gradient-free optimization platform. https://GitHub.com/FacebookResearch/Nevergrad.