Improving and extending Probabilistic Regression trees

Machine learning methods have been successful in various domains, such as marketing with customer behaviour prediction, health with patient diagnosis and industry with the optimisation of industrial processes. In many cases, one needs to make some prediction from parameters that are heterogeneous, as they can be quantitative or qualitative, ordinal or non-ordinal, real or Boolean, and are uncertain in the sense that their values usually originate from noisy measurement procedures. In practice, it is important to combine all these heterogeneous parameters and to take into account their uncertainty to improve the predictions made.

To address these issues, we have recently introduced [1] a new model called Probabilistic Regression (PR) trees that extend standard regression trees with the possibility to adapt to the smoothness of the prediction function while preserving interpretability and being robust to noise. This project is intended to further develop this model and make the current research prototypes more robust. In particular, the successful candidate will have to address the following points:

- Evaluate the quality of the prediction made by PR trees on several real and challenging datasets as well as the impact of uncertainty on the results;
- Design, implement and test new machine learning/data analysis methods to *e.g.* assess the advantages/disadvantages of the quantile version of PR trees and determine the importance of each parameter;
- Make the current research prototypes more usable and robust.

Context This project fits within the Grenoble Computer Science Lab (called LIG, http://www.liglab.fr/en) and the Interdisciplinary Institute in Artificial Intelligence MIAI@Grenoble Alpes (https://miai.univ-grenoble-alpes.fr/). MIAI@Grenoble Alpes is one of the four AI Institutes created by the French government to accelerate R&D, teaching and innovation in AI in France. It is also based on a collaboration with Marianne Clausel in IECL (Nancy) and with two industrial partners, namely Total and Serimax.

To apply Interested candidates should send a complete CV with a list of publications and two reference letters to Emilie Devijver (emilie.devijver@univ-grenoble-alpes.fr) and Eric Gaussier (eric.gaussier@univ-grenoble-alpes.fr). Candidates should have excellent software engineering skills. They should also have experience in machine learning and modelling, an ability to work effectively with a multidisciplinary team of computer scientists and mathematicians, and excellent oral and written communication skills.

Starting date and duration The postdoc is intended for 18 months, starting as soon as possible and no later than June 2021.

Location The work should take place on the University Campus in Grenoble, France.

References

[1] Sami Alkhoury, Emilie Devijver, Marianne Clausel, Myriam Tami, Eric Gaussier and Georges Oppenheim. *Smooth And Consistent Probabilistic Regression Trees*. Advances in Neural Information Processing Systems 33 pre-proceedings (NeurIPS 2020)