Efficient Interpretable Machine Learning with Truly Unordered Rule Sets

Lincen Yang

1 Background

Intrinsically interpretable machine learning models such as decision trees and (probabilistic) classification rules have received a lot of interest recently, due to the need for deploying machine learning systems in critical regions as well as for knowledge discovery from data.

We consider the truly unordered rule set (TURS) model,[1] arguably the most comprehensible rule-based model for domain experts, which have been demonstrated both to have simple model complexity (in terms of the number of rules and rule lengths) and to be fundamentally simpler when a single prediction needs to be explained.

This project aims to further strengthen the method for learning a TURS model from data from various perspectives. We are looking for 2 - 3 master students who are self-motivated and intrinsically interested in developing cutting-edge interpretable machine learning methods for tabular data.

2 Projects

Project I: Software development for TURS. This project has two aims. The first aim is to improve the current implementation of TURS in Python, using advanced tools such as Numba and Cython to improve the efficiency of the code. The second aim is to integrate the implementation to existing Python packages like 'imodels' (`https://github.com/csinva/imodels`).

Project II: Fast approximations for TURS. This project aims to develop a highly efficient probabilistic model to approximate TURS but meanwhile maintain its performance (as much as possible). The main challenge is how to keep the property of being "truly unordered".

Project III: Benchmark against neural-symbolic rule learning methods. This project aims to understand from various perspectives the performance

[1] Yang, Lincen, and Matthijs van Leeuwen. "Truly unordered probabilistic rule sets for multi-class classification." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2022.

of TURS in comparison with the very recently proposed neural-network-based rule learning methods, by running large-scale benchmark experiments.

3 Requirements

We are looking for students with strong programming skills in Python. Familiarity with basic machine learning concepts and Probability & Statistics knowledge is required. Previous experience with academic writing and/or knowledge of information theory and the MDL principle is a plus.

Interested students can send your CV to `l.yang@liacs.leidenuniv.nl`, together with a short motivation letter (up to half a page) to explain 1) which of the projects you would like to work on, and 2) how your background & skills match the project.