RL-Net: Interpretable Rule Learning with Neural Networks

Lucile Dierckx[1,2,*], Rosana Veroneze[2,3] and Siegfried Nijssen[1,2]

[1]TRAIL Institute, Louvain-la-Neuve, Belgium
[2]ICTEAM/INGI, UCLouvain, Louvain-la-Neuve, Belgium
[3]FEEC/DCA, Unicamp, Campinas-SP, Brazil

Keywords
Interpretability, Pattern set mining, Rule learning, Binary neural networks

This paper was accepted for publication in the proceedings of PAKDD 2023 [1].

Many areas in which classifiers are applied do not only require high predictive performance, but also interpretability of the predictions. An important category of interpretable symbolic classifiers are rule-based models. Most rule-based algorithms use heuristics to learn the models and are designed for specific classification tasks.

Recent studies have explored the use of gradient-based approaches from the neural network literature to learn this type of classifier, combining the strengths of both neural and symbolic learning to leverage the literature of neural networks for rule learning. However, these studies focused on (non-ordered) rule sets, while much of the literature on rule learning uses (ordered) rule lists. Rule sets are based on a `IF ... THEN Positive; IF ... THEN Positive; (...); ELSE Negative` structure, where every rule predicts the same class, while rule list follow a `IF ... THEN ...; ELSE IF ... THEN ...; (...); ELSE ...` pattern, where every rule can predict a different class. An interesting aspect of rule lists is that they remain fully interpretable for multi-class classification tasks.

In this work, we extended Qiao et al.'s neural network for learning rule sets [2] by adding hierarchy among the rules, which allows us to learn rule lists. Additionally, our model can solve multi-class classification problems instead of being limited to binary problems. Finally, our proposal can be easily adapted to solve multi-label classification problems. Overall, our work provides a novel contribution to the field of neuro-symbolic learning by introducing an approach for learning neural networks that represent interpretable classification rules.

Our neural network for rule learning (RL-Net) utilizes the neural network's structure and its gradient-based learning optimization to learn rule lists. It is composed of four layers, as depicted in Figure 1. The first one receives the dataset's binary features. The input is then connected to the rule layer that learns the conditions composing the different rules. The next

NeSy 2023, 17th International Workshop on Neural-Symbolic Learning and Reasoning, Certosa di Pontignano, Siena, Italy
*Corresponding author.
lucile.dierckx@uclouvain.be (L. Dierckx); rosana.veronze@uclouvain.be (R. Veroneze); siegfried.nijssen@uclouvain.be (S. Nijssen)
0000-0003-2855-1042 (L. Dierckx); 0000-0003-4007-9350 (R. Veroneze); 0000-0003-2678-1266 (S. Nijssen)

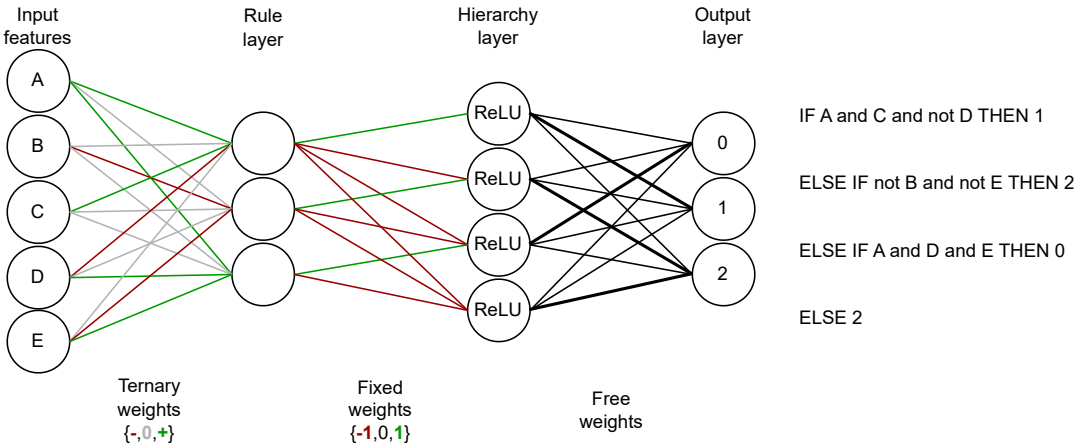

Figure 1: Architecture of RL-Net and its equivalent translation.

layer expresses the hierarchy among the rules, which is necessary to learn rule lists instead of rule sets. Finally, the output layer gives a class label to each rule. To remain fully interpretable, the activations and weights of the model's nodes are mostly binary or ternary in order to mimic the behavior of the logical operation AND and the hierarchy between the rules. The four different layers are described in detail in the original paper and the source code is available online[1].

RL-Net performs similarly to the state-of-the-art rule-based algorithms on binary and multi-class classification problems. Moreover, we showed that RL-Net can easily be adapted to multi-label learning tasks but cannot fully compete with the widely used algorithms yet. A presentation and discussion of the experiments is provided in the original paper.

Further research includes working on the multi-label classification, taking better advantage of the maximal number of rules that are allowed by the user, and integrating our RL-Net in some developed fields of the neural network literature.

Acknowledgments

This work was supported by Service Public de Wallonie Recherche under grant n°2010235 – ARIAC by DIGITALWALLONIA4.AI, and n°2110107 - SERENITY2 by WIN2WAL. We also thank FAPESP, Brazil (Grants No. 2017/21174-8 and 2020/00123-9) for the financial support.

References

[1] L. Dierckx, R. Veroneze, S. Nijssen, Rl-net: Interpretable rule learning with neural networks, in: PAKDD 2023: Advances in Knowledge Discovery and Data Mining, Springer International Publishing, 2023.

[2] L. Qiao, W. Wang, B. Lin, Learning accurate and interpretable decision rule sets from neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2021.

[1] https://github.com/luciledierckx/RLNet