RL-Net: Apprentissage de Règles Interprétables avec des Réseaux de Neurones

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

[1] Institut TRAIL, Louvain-la-Neuve, Belgique
[2] UCLouvain, ICTEAM/INGI, Louvain-la-Neuve, Belgique
[3] Unicamp, FEEC/DCA, Campinas-SP, Brésil

{prénom.nom}@uclouvain.be

Résumé

Les modèles symboliques et interprétables, tels que ceux à règles, sont étudiés depuis longtemps car de nombreux domaines ont besoin d'interpétabilité. De récents travaux ont étudié l'apprentissage de règles utilisant des approches pour apprendre des réseaux neuronaux plutôt que des heuristiques. Ces travaux se limitent aux règles non-ordonnées. Nous proposons RL-Net, qui apprend des règles ordonnées avec un réseau neuronal. Notre modèle performe de manière similaire aux algorithmes de pointe pour la classification binaire et multiclasse, et peut être adapté aux tâches à étiquettes multiples.

Mots-clés

Interprétabilité, extraction d'ensembles de motifs, apprentissage de règles, réseaux neuronaux binaires.

Abstract

Symbolic, interpretable models, such as rule-based models, have been studied for years due to the need for interpretable classification models in many domains. Recent studies have explored gradient-based rule learning using neural networks instead of heuristics. However, these works focus on unordered rule sets only. We propose RL-Net, an approach for learning ordered rule lists based on neural networks. Our model performs similarly to state-of-the-art algorithms for rule learning in binary and multi-class classification settings, and can be adapted to multi-label tasks.

Keywords

Interpretability, pattern set mining, rule learning, binary neural networks.

1 Introduction

Cet article a été accepté à PAKDD 2023 [6]

De nombreux domaines d'application requièrent des modèles de classification non seulement performants mais aussi interprétables. C'est pourquoi, la recherche sur la classification symbolique et interprétable est un domaine fort présent dans la littérature. Une catégorie importante de ces classificateurs interprétables est celle des modèles basés sur des règles de logique [4, 5]. Ces classifieurs à base de règles utilisent des heuristiques pour apprendre les règles interprétables et sont conçus pour des tâches de classification spécifiques. De récentes études se sont penchées sur l'utilisation d'approches basées sur le gradient des réseaux neuronaux afin d'apprendre ce type de classificateur. Leur objectif est de combiner l'apprentissage neuronal et l'apprentissage symbolique pour pouvoir ensuite exploiter la littérature sur les réseaux neuronaux dans le cadre de l'apprentissage de règles. Les articles existants sur cette combinaison neuro-symbolique se concentrent sur les classificateurs utilisant des règles non-ordonnées [2, 3, 7, 9, 10, 11]. Cependant, une grande partie de la littérature portant sur les modèles à règles utilise des règles ordonnées [5], qui ont l'avantage de rester entièrement interprétables, même pour des tâches de classification avec plus de deux classes. La classification avec des listes de règles ordonnées n'a pas encore été étudiée dans un cadre neuro-symbolique entièrement interprétable.

Dans ce travail, nous nous concentrons sur cette classification en étendant le réseau de règles de décision (DR-Net) de Qiao et al [9]. Nous avons implémenté une hiérarchie entre les règles, permettant ainsi d'apprendre des classificateurs basés sur des listes ordonnées de règles au lieu d'ensembles non-ordonnés de règles. De plus, notre modèle permet de résoudre des problèmes de classification multiclasses au lieu d'être limité aux problèmes binaires. Enfin, notre proposition peut facilement être adaptée pour résoudre les problèmes de classification à étiquettes multiples.

2 Approche

Notre réseau de neurones pour l'apprentissage de règles (RL-Net) a été conçu pour apprendre des listes de règles interprétables qui peuvent faire de la classification multiclasse. RL-Net utilise la structure d'un réseau neuronal ainsi que son apprentissage par optimisation du gradient.

Le réseau neuronal que nous avons imaginé reproduit le modèle `SI ... ALORS ... ; SINON SI ... ALORS ... ; (...) ; ENFIN ...`. Pour ce faire, il est composé de quatre couches, tel qu'illustré Figure 1. La première couche est la couche d'entrée qui reçoit les carac-

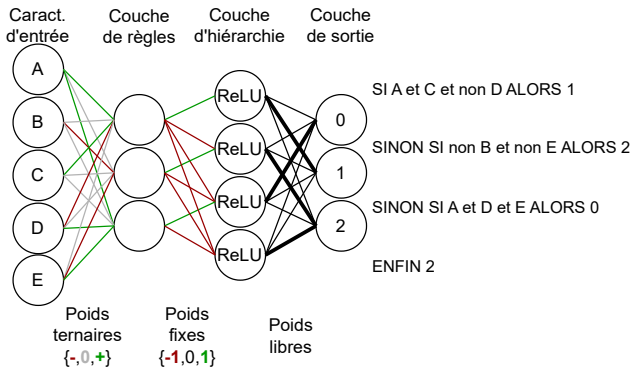

FIGURE 1 – Architecture de RL-Net

téristiques binarisées du jeu de données. Elle est connectée à la couche de règles, où les conditions composant les différentes règles sont apprises. La couche suivante exprime la hiérarchie [1] entre les règles, ce qui est nécessaire pour apprendre une liste de règles ordonnée au lieu d'un ensemble de règles non-ordonné. Enfin, la couche de sortie attribue une étiquette de classe spécifique à chaque règle. Pour rester entièrement interprétable, les activations et les poids des noeuds du modèle sont tous binarisés ou ternarisés [8] de façon à imiter le comportement de l'opération logique `ET` et la hiérarchie entre les règles. Grâce à son architecture neuronale, l'apprentissage de notre modèle de règles ordonnées peut être fait à l'aide des techniques classiques d'apprentissage des réseaux de neurones (telles que l'optimiseur Adam, la minimisation de l'entropie, ...). Une description plus détaillée de chaque couche du modèle est faite dans l'article original. Le code source se trouve sur GitHub [1].

3 Résultats

Le modèle que nous avons développé peut être utilisé pour des tâches de classification binaire et multiclasse, et atteint des performances similaires aux algorithmes fréquemment utilisés, RIPPER [5] et CART [4], en particulier lorsque le nombre maximal de règles défini par l'utilisateur est peu élevé. Nous avons également travaillé sur l'adaptation de notre modèle au contexte à étiquettes multiples. Dans cette configuration, nous observons que le modèle possède un bon potentiel d'apprentissage, mais ne peut pas encore rivaliser avec les algorithmes bien connus.
La description des datasets et expériences ainsi que les résultats détaillés sont disponibles dans l'article original.

4 Conclusion

Nous avons proposé RL-Net, une approche par réseau de neurones pour l'apprentissage de listes de règles ordonnées. Notre modèle a donné des résultats similaires à ceux des algorithmes de pointe pour l'apprentissage de règles dans des contextes de classification binaire et multiclasse, et a du potentiel pour être adapté aux tâches d'apprentissage à étiquettes multiples. Les pistes d'amélioration pour le modèle seraient de travailler plus en profondeur sur l'adaptation à la classification à étiquettes multiples ainsi que d'arriver à prendre plus avantage du nombre maximal de règles que l'utilisateur autorise le modèle à utiliser.

1. `https://github.com/luciledierckx/RLNet`

Remerciements

Ce travail a été soutenu par le Service Public de Wallonie Recherche sous la subvention n°2010235 - ARIAC by DIGITALWALLONIA4.AI et n°2110107 - SERENITY2 by WIN2WAL. R. Veroneze tient aussi à remercier FAPESP, Brésil (Subventions n°2017/21174-8 et 2020/00123-9).

Références

[1] John OR Aoga, Siegfried Nijssen, and Pierre Schaus. Modeling pattern set mining using boolean circuits. In *International Conference on Principles and Practice of Constraint Programming*. Springer, 2019.

[2] Florian Beck and Johannes Fürnkranz. An empirical investigation into deep and shallow rule learning. *Frontiers in Artificial Intelligence*, 4, 2021.

[3] Florian Beck and Johannes Fürnkranz. An investigation into mini-batch rule learning. *arXiv preprint arXiv :2106.10202*, 2021.

[4] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.

[5] William W Cohen. Fast effective rule induction. In *Twelfth International Conference on Machine Learning*, pages 115–123. Elsevier, 1995.

[6] Lucile Dierckx, Rosana Veroneze, and Siegfried Nijssen. Rl-net : Interpretable rule learning with neural networks. In *PAKDD 2023 : Advances in Knowledge Discovery and Data Mining*. Springer International Publishing, 2023.

[7] Jonas Fischer and Jilles Vreeken. Differentiable pattern set mining. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 383–392, 2021.

[8] Christos Louizos, Max Welling, and Diederik P. Kingma. Learning sparse neural networks through L0 regularization. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.

[9] Litao Qiao, Weijia Wang, and Bill Lin. Learning accurate and interpretable decision rule sets from neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4303–4311, 2021.

[10] Shaoyun Shi, Yuexiang Xie, Zhen Wang, Bolin Ding, Yaliang Li, and Min Zhang. Explainable Neural Rule Learning. In *WWW 2022 - Proceedings of the ACM Web Conference 2022*, pages 3031–3041. Association for Computing Machinery, Inc, 2022.

[11] Yongxin Yang, Irene Garcia Morillo, and Timothy M Hospedales. Deep neural decision trees. *ICML Workshop on Human Interpretability in Machine Learning. arXiv preprint arXiv :1806.06988*, 2018.