Set-valued prediction in hierarchical classification

Thomas Mortier[1], Eyke Hüllermeier[2], Krzysztof Dembczyński[3][4], and Willem Waegeman[1]

[1] Dept. of Data Analysis and Mathematical Modelling, Ghent University, Belgium
[2] Institute of Informatics, LMU Munich, Germany
[3] Institute of Computing Science, Poznań University of Technology, Poland
[4] Yahoo! Research, New York, USA

Abstract Set-valued prediction is a well-known concept in multi-class classification. When a classifier is uncertain about the class label for a test instance, it can predict a set of classes instead of a single class. In this paper, we focus on hierarchical multi-class classification problems, where valid sets (typically) correspond to internal nodes of the hierarchy. We argue that this is a very strong restriction, and we propose a relaxation by introducing the notion of representation complexity for a predicted set. In combination with probabilistic classifiers, this leads to a challenging inference problem for which specific combinatorial optimization algorithms are needed. We propose three methods and evaluate them on benchmark datasets: a naïve approach that is based on matrix-vector multiplication, a reformulation as a knapsack problem with conflict graph, and a recursive tree search method. Experimental results demonstrate that the last method is computationally more efficient than the other two approaches, due to a hierarchical factorization of the conditional class distribution.

Keywords: Multi-class classification · Hierarchical classification · Set-valued prediction.

1 Introduction

In multi-class classification problems with a lot of classes, there are often situations where a classifier is uncertain about the class label for a given instance, e.g., because of class ambiguity. Set-valued predictions form a natural way of dealing with this uncertainty, by predicting a set of classes instead of a single class. In the machine learning literature, set-valued prediction has been studied under different frameworks. A simple approach consists of top-k prediction, i.e., returning a set with the k classes that have the highest probabilities or scores [7,2]. Another popular approach is conformal prediction [12], which produces sets that contain the true class with high probability. A third framework is rooted in Bayesian decision theory and optimizes a utility function that trades off two important criteria for set-valued predictions, namely correctness and precision [5,3,4,15,14,9]. Set-valued prediction has also been considered in a hierarchical classification setting, where similarity among classes is encoded by means of a predefined class hierarchy provided by domain experts. In hierarchical classification, set-valued predictions are often restricted to specific subsets of the set of classes, namely those that correspond to nodes of the hierarchy and, therefore, have a clear interpretation and are deemed semantically meaningful [6,1,11,13]. On the other side, a restriction of that kind may negatively impact predictive performance. That's why a few authors allow any subset of classes as a prediction in hierarchical classification – ignoring the hierarchy altogether – which comes at the expense of semantic complexity [10,9].

2 Proposed framework

We propose a novel set-valued prediction framework for hierarchical classification that makes a compromise between the two extremes [8]. Compared to approaches that predict a single node of the hierarchy, we will be less restrictive in the type of sets that can be returned, but we will be more restrictive than methods that return any subset of classes. To this end, we present a decision-theoretic framework with an inference procedure at prediction time, where we aim to find the set with highest probability mass, while restricting the so-called representation complexity and set size. We propose three different methods that solve the inference problem in an exact way: a naïve approach that is based on matrix-vector multiplication, a reformulation as a knapsack problem with conflict graph, and a recursive tree search method. We illustrate the usefulness of our proposed framework on a fine-grained visual categorization dataset and compare the different algorithms in terms of predictive performance and runtime efficiency for five different benchmark datasets. Experimental results demonstrate that the last method is computationally more efficient than the other two approaches, due to a hierarchical factorization of the conditional class distribution.

3 Conclusion

We briefly discussed a new decision-theoretic framework for set-valued prediction in hierarchical classification by introducing the notion of representation complexity. This complexity allows the user to relax the often strong restriction that is implied by hierarchical classification, namely that predictions should correspond to single nodes of a predefined hierarchy. Several algorithms are proposed that solve the challenging optimization problem in an exact way. One of those algorithms, based on a recursive tree search method that uses a hierarchical factorization of the conditional class distribution, shows especially promising results in terms of runtime complexity.

References

1. Bi, W., Kwok, J.: Bayes-optimal hierarchical multilabel classification. IEEE Trans. on Know. and Data Eng. **27** (11 2015)
2. Chzhen, E., Denis, C., Hebiri, M., Lorieul, T.: Set-valued classification - overview via a unified framework. ArXiv (2021). https://doi.org/abs/2102.12318
3. Corani, G., Zaffalon, M.: Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2. JMLR **9**, 581–621 (2008)
4. Corani, G., Zaffalon, M.: Lazy naive credal classifier. In: Proc. of the 1st ACM SIGKDD Workshop on Knowledge Discovery from Uncertain Data. pp. 30–37 (2009)
5. Del Coz, J.J., Díez, J., Bahamonde, A.: Learning nondeterministic classifiers. JMLR **10**, 2273–2293 (2009)
6. Freitas, A., de Carvalho, A.: A tutorial on hierarchical classification with applications in bioinformatics. In: Research and Trends in Data Mining Technologies and Applications,. pp. 175–208 (2007)
7. Lapin, M., Hein, M., Schiele, B.: Loss functions for top-k error: Analysis and insights. In: CVPR. pp. 1468–1477 (2016)
8. Mortier, T., Hüllermeier, E., Dembczyński, K., Waegeman, W.: Set-valued prediction in hierarchical classification with constrained representation complexity. In: Cussens, J., Zhang, K. (eds.) Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence. Proceedings of Machine Learning Research, vol. 180, pp. 1392–1401. PMLR (01–05 Aug 2022)
9. Mortier, T., Wydmuch, M., Dembczyński, K., Waegeman, W.: Efficient set-valued prediction in multi-class classification. KDD **35**, 1435–1469 (2021)
10. Oh, S.: Top-k hierarchical classification. In: AAAI. pp. 2450–2456. AAAI Press (2017)
11. Rangwala, H., Naik, A.: Large scale hierarchical classification: foundations, algorithms and applications. In: The European Conf. on ML and Principles and Practice of Know. Disc. in Datab. (2017)
12. Shafer, G., Vovk, V.: A tutorial on conformal prediction. JMLR **9**, 371–421 (2008)
13. Yang, G., Destercke, S., Masson, M.H.: Cautious classification with nested dichotomies and imprecise probabilities. Soft Computing **21**, 7447–7462 (2017)
14. Yang, G., Destercke, S., Masson, M.H.: The costs of indeterminacy: How to determine them? IEEE Transactions on Cybernetics **47**, 4316–4327 (2017)
15. Zaffalon, M., Giorgio, C., Mauá, D.D.: Evaluating credal classifiers by utility-discounted predictive accuracy. Int. J. Approx. Reasoning **53**, 1282–1301 (2012)