Reliable Multilabel Classification: Prediction with Partial Abstention *

Vu-Linh Nguyen VU.LINH.NGUYEN@UNI-PADERBORN.DE
Eyke Hüllermeier EYKE@UPB.DE
Heinz Nixdorf Institute and Department of Computer Science, Paderborn University, Germany

In statistics and machine learning, classification with abstention, also known as classification with a reject option, is an extension of the standard setting of classification, in which the learner is allowed to refuse a prediction for a given query instance [2]. For the learner, the main reason to abstain is a lack of certainty about the corresponding outcome, i.e., refusing or at least deferring a decision might then be better than taking a high risk of a wrong decision.

Nowadays, there are many machine learning problems in which complex, structured predictions are sought (instead of scalar values, like in classification and regression). For such problems, the idea of abstaining from a prediction can be generalized toward *partial abstention*: Instead of predicting the entire structure, the learner predicts only parts of it, namely those for which it is certain enough. This idea has already been realized, for example, for the problem of *label ranking*, where predictions are rankings [1].

Another important example is *multi-label classification* (MLC), in which an outcome associated with an instance is a labeling in the form of a subset of an underlying reference set of class labels; that is, the output space is the power set of that reference set [3]. The MLC problem can be formalized in form of a risk minimization problem [3]. Given an underlying MLC loss, e.g, Hamming or $0/1$ subset loss, and assuming a probabilistic prediction for a query instance on the set of labelings (or at least an estimation thereof), the problem of risk minimization comes down to finding the optimal prediction, i.e., the labeling which minimizes this MLC loss in expectation. The concrete form of this optimization problem as well as its difficulty depend on several choices, including the underlying MLC loss function.

In this paper, we study an extension of the setting of MLC, in which the learner is allowed to partially abstain from a prediction, that is, to deliver predictions on some but not necessarily all class labels (or, more generally, to refuse committing to a single complete prediction). Although MLC has been studied extensively in the machine learning literature in the recent past, there is surprisingly little work on MLC with abstention so far. A notable exception is an approach by Pillai et al. [4], whose main idea is discussed in the poster.

We present a formal framework of MLC with partial abstention, which builds on two main building blocks: First, the extension of an underlying MLC loss function so as to accommodate abstention in a proper way, and second the problem of finding the *optimal prediction with partial abstention*, that is, minimizing this extended loss in expectation. We discuss some general properties that might be of interest for extended losses and propose a generic formulation of extended losses based on an additive penalty. We instantiated our framework for the case of the Hamming and rank loss, both theoretical and experimental. In both cases, we elaborated on properties of optimal predictions, showed them to have a specific structure, and devised efficient methods—in time $O(m\log(m))$, when m is the number of the labels—to produce optimal predictions. Experimentally, we showed these methods to be effective in the sense of reducing loss when being allowed to abstain.

References

[1] Weiwei Cheng, Eyke Hüllermeier, Willem Waegeman, and Volkmar Welker. Label ranking with partial abstention based on thresholded probabilistic models. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2501–2509, 2012.

[2] Corinna Cortes, Giulia DeSalvo, and Mehryar Mohri. Learning with rejection. In *Proceedings of the 27th International Conference on Algorithmic Learning Theory (ALT)*, pages 67–82. Springer Verlag, 2016.

[3] Krzysztof Dembczyński, Willem Waegeman, Weiwei Cheng, and Eyke Hüllermeier. On label dependence and loss minimization in multi-label classification. *Machine Learning*, 88(1-2):5–45, 2012.

[4] Ignazio Pillai, Giorgio Fumera, and Fabio Roli. Multi-label classification with a reject option. *Pattern Recognition*, 46 (8):2256–2266, 2013.

* *Acknowledgments* We thank Willem Waegeman for interesting discussions and useful suggestions. This work has been supported by the Germany Research Foundation (DFG)