Université de technologie de Compiègne – PhD thesis proposal

Robust ensembling in set-valued classification

Application deadline: May 03, 2026
Duration: 3 years starting from October 2026
Grant: Allocation MESR (https://www.enseignementsup-recherche.gouv.fr/fr/le-financement-doctoral-46472**)**
Place : Université de technologie de Compiègne – Heudiasyc, 60200 Compiègne, France

Supervisors:
- Thierry Denoeux, Professor at UTC and senior member of IUF (thierry.denoeux@hds.utc.fr)
- Vu-Linh Nguyen, Junior professor (vu-linh.nguyen@hds.utc.fr)

This PhD thesis proposal falls within the scope of the research group CID of the Heudiasyc laboratory of the Université de technologie de Compiègne. The primary focus is to develop a new framework for robust ensembling in set-valued classification. The selected candidate will have the opportunity to work in an interdisciplinary environment in collaboration with active researchers in the Netherlands, UK, France, and other countries.

Scientific context

Ensemble learning is a common practice to achieve robustness and improved generalization in machine learning [7]. In classification tasks, explored in this thesis project, ensemble learning typically consists of two phases: *Generation* and *aggregation*. The former trains a set of predictive models that seek a diversity-accurateness trade-off. During the latter, for each instance, the predictions of predictive models are aggregated into the final prediction.

Although conventional ensemble learning is advantageous in various application domains, the assumption that predictive models always produce singleton predictions, which should be, in turn, aggregated into a final singleton prediction, can lead to challenging problems in the presence of *low coverage* in ensemble aggregation. By low coverage, we refer to the scenario in which, for the current instance, only a minority or none of the predictive models provide the correct prediction. In fact, low coverage is difficult to handle given the voting nature of the aggregation phase [7], and may lead to an unsatisfactory correctness level of the final predictions.

Low coverage is an issue even in the simple *multiclass classification (MCC)* setting, where one has to predict a single class variable. This issue is significantly amplified in the following more challenging classification tasks that require large data sets: *multilabel classification (MLC)*, that is, predicting multiple binary variables simultaneously [5] and *multidimensional classification (MDC)*, that is, predicting multiple categorical variables simultaneously [6].

Objectives

We aim to construct a new ensemble learning framework, especially designed to tackle the low coverage issue, aiming at both coverage and diversity:

- ***Generation***: Train a set of set-valued classifiers [4, 5, 7] that seek both coverage, i.e., the true outcome is included in the set-valued predictions, and diversity.
- ***Aggregation***: For each instance, the set-valued predictions of predictive models are aggregated into the final prediction, which can be either a single outcome or a set of outcomes.

We shall focus on addressing the two crucial questions:

- Can we eventually achieve both coverage and diversity?
- If so, how could these improvements translate into an enhanced predictive performance of the ensemble?

We plan to generalize contributions on combining partial information following the evidential approach [1, 2, 3], and set-valued classification [4, 5, 7] to seek theoretically sound answers to these questions. We also plan to focus on the mathematical aspects of optimization problems to gain the scalability of the framework.

Required profile

- M2 or engineering diploma specialized in computer science
- Good programming skills (Python, PyTorch, TensorFlow, ...) and/or a strong background in mathematics.
- Autonomy, rigor and critical thinking.

Application

If you are interested in this position, please send the following documents to Prof. Thierry Denoeux (thierry.denoeux@hds.utc.fr) and Vu-Linh Nguyen (vu-linh.nguyen@hds.utc.fr):

- ***CV***
- ***Motivation letter (explaining why this specific position interests you)***
- ***Academic transcripts from your Bachelor's, Master's or Engineering degree***
- ***Name and contact information of at least two references we may contact if necessary***

References

[1] T. Denoeux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE transactions on systems, man, and cybernetics, 25(5):804–813, 1995.

[2] T. Denoeux. A neural network classifier based on dempster-shafer theory. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 30(2): 131–150, 2000.

[3] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial Intelligence, 172(2-3):234–264, 2008.

[4] L. Ma and T. Denoeux. Making set-valued predictions in evidential classification: A comparison of different approaches. In ISIPTA, pages 276–285, 2019.

[5] V.-L. Nguyen and E. H¨ullermeier. Multilabel classification with partial abstention: Bayes-optimal prediction under label independence. Journal of Artificial Intelligence Research, 72:613–665, 2021.

[6] V.-L. Nguyen, Y. Yang, and C. P. de Campos. Probabilistic multi-dimensional classification. In UAI, pages 1522–1533, 2023.

[7] V.-L. Nguyen, H. Zhang, and S. Destercke. Credal ensembling in multi-class classification. Machine Learning, 114(1):1–62, 2025.