

## Lazy Credal Classifier and how to compare credal classifiers

**Giorgio Corani**  
IDSIA, Switzerland  
giorgio@idsia.ch

**Marco Zaffalon**  
IDSIA, Switzerland  
zaffalon@idsia.ch

### Abstract

This poster carries out two main contributions: (a) a lazy (or local) version of naive credal classifier (NCC) that we call *lazy naive credal classifier* (LNCC); (b) two metrics to compare credal classifiers.

NCC [1] has extended *naive Bayes* (NB) to imprecise probabilities, by modeling prior ignorance via the Imprecise Dirichlet Model; the classification is eventually issued by returning the *non-dominated* classes; therefore, NCC returns a set of classes when faced with instances whose classification would be prior-dependent for NB. Extensive experiments have shown that NCC is more reliable than NB. Yet, two drawbacks of NCC are (i) that the naive assumption (statistical independence of the features given the class) might be too simplistic and (ii) that in some cases NCC becomes too indeterminate.

We address the two issues above proposing the *local naive credal classifier* (LNCC); this addresses point (i) because working locally reduces the chance of encountering strong dependencies. In addition, LNCC should improve also the determinacy of NCC, thus addressing (ii): by working locally, it selects the part of the learning set that is more informative about the instance to be classified. How do we select the number of instances to be used to train the local classifier? We keep on including instances in the local learning set until NCC starts issuing a determinate classification on the instance to classify (note that this clearly favors removing indeterminacy). The rationale behind this criterion is that we select a local learning set that is informative enough to draw a strong conclusion, such as a determinate classification.

We investigate the effect of the above choices by extensive experiments to compare LNCC with NCC. In order to compare LNCC and NCC, we propose (a) an indicator borrowed from multi-label classification and (b) a non-parametric rank test. To our knowledge, this is the first attempt to empirically compare credal classifiers.

Results on 36 data sets show that, according to both tests, LNCC clearly outperforms NCC, as it significantly reduces indeterminacy without worsening (often improving) the overall accuracy.

### Acknowledgments

Work partially supported by the Swiss NSF grant n. 200021-118071/1.

### References

- [1] G. Corani and M. Zaffalon. Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2. *Journal of Machine Learning Research*, 9:581–621, 2008.