

Putting Things in Order: On the Fundamental Role of Ranking in Classification and Probability Estimation*

Peter A. Flach

Department of Computer Science, University of Bristol, United Kingdom
Peter.Flach@bristol.ac.uk

Abstract. While a binary classifier aims to distinguish positives from negatives, a ranker orders instances from high to low expectation that the instance is positive. Most classification models in machine learning output some score of ‘positiveness’, and hence can be used as rankers. Conversely, any ranker can be turned into a classifier if we have some instance-independent means of splitting the ranking into positive and negative segments. This could be a fixed score threshold; a point obtained from fixing the slope on the ROC curve; the break-even point between true positive and true negative rates; to mention just a few possibilities.

These connections between ranking and classification notwithstanding, there are considerable differences as well. Classification performance on n examples is measured by accuracy, an $O(n)$ operation; ranking performance, on the other hand, is measured by the area under the ROC curve (AUC), an $O(n \log n)$ operation. The model with the highest AUC does not necessarily dominate all other models, and thus it is possible that another model would achieve a higher accuracy for certain operating conditions, even if its AUC is lower.

However, within certain model classes good ranking performance and good classification performance are more closely related than suggested by the previous remarks. For instance, there is evidence that certain classification models, while designed to optimise accuracy, in effect optimise an AUC-based loss function [1]. It has also been known for some time that decision tree yield convex training set ROC curves by construction [2], and thus optimising training set accuracy is likely to lead to good training set AUC. In this talk I will investigate the relation between ranking and classification more closely.

I will also consider the connection between ranking and probability estimation. The quality of probability estimates can be measured by, e.g., mean squared error in the probability estimates (the Brier score). However, like accuracy, this is an $O(n)$ operation that doesn’t fully take ranking performance into account. I will show how a novel decomposition of the Brier score into calibration loss and refinement loss [3] sheds light on both ranking and probability estimation performance. While previous decompositions are approximate [4], our decomposition is an exact one based on the ROC convex hull. (The connection between the ROC convex hull and calibration was independently noted by [5]). In the case of decision trees, the analysis explains the empirical evidence that probability estimation trees produce well-calibrated probabilities [6].

* Invited speakers at ECML/PKDD are supported by the PASCAL European network of excellence.

References

1. Rudin, C., Cortes, C., Mohri, M., Schapire, R.E.: Margin-based ranking meets boosting in the middle. In: Auer, P., Meir, R. (eds.) COLT 2005. LNCS (LNAI), vol. 3559, pp. 63–78. Springer, Heidelberg (2005)
2. Ferri, C., Flach, P.A., Hernández-Orallo, J.: Learning decision trees using the area under the ROC curve. In: Proceedings of the Nineteenth International Conference on Machine Learning (ICML 2002), pp. 139–146. Morgan Kaufmann, San Francisco (2002)
3. Flach, P.A., Matsubara, E.T.: A simple lexicographic ranker and probability estimator. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) ECML 2007. LNCS, vol. 4701, Springer, Heidelberg (2007)
4. Cohen, I., Goldszmidt, M.: Properties and benefits of calibrated classifiers. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) PKDD 2004. LNCS (LNAI), vol. 3202, pp. 125–136. Springer, Heidelberg (2004)
5. Fawcett, T., Niculescu-Mizil, A.: PAV and the ROC convex hull. *Machine Learning* 68(1), 97–106 (2007)
6. Niculescu-Mizil, A., Caruana, R.: Predicting good probabilities with supervised learning. In: Proceedings of the Twenty-Second International Conference on Machine Learning (ICML 2005), pp. 625–632. ACM, New York (2005)