

A Fast Ensemble Pruning Algorithm Based on Pattern Mining Process*

Qiang-Li Zhao, Yan-Huang Jiang, and Ming Xu

School of Computer Science, National University of Defense Technology,
Changsha, Hunan Province, PR. China
zhao-qiangli@163.com, yhjiang@nudt.edu.cn

Abstract. Ensemble pruning deals with the reduction of base classifiers prior to combination in order to improve generalization and prediction efficiency. Existing ensemble pruning algorithms require much pruning time. This paper presents a fast pruning approach: PMEP (Pattern Mining based Ensemble Pruning). In this algorithm, the prediction results of all base classifiers are organized as a transaction database, and FP-Tree structure is used to compact the prediction results. Then a greedy pattern mining method is explored to find the ensemble of size k . After obtaining the ensembles of all possible sizes, the one with the best accuracy is outputted. Compared with Bagging, GASEN, and Forward Selection, experimental results show that PMEP achieves the best prediction accuracy and keeps the size of the final ensemble small, more importantly, its pruning time is much less than other ensemble pruning algorithms.

Keywords: PMEP (Pattern Mining based Ensemble Pruning), FP-Tree, Bagging, back-propagation neural network.

Reference

1. Zhao, Q.-L., Jiang, Y.-H., Xu, M.: A Fast Ensemble Pruning Algorithm based on Pattern Mining Process. *Data Mining and Knowledge Discovery* (2009) DOI: 10.1007/s10618-009-0138-1

* This is an extended abstract of an article published in the *Data Mining and Knowledge Discovery* journal [1].