

Lazy Learning for Improving Ranking of Decision Trees

Han Liang, Yuhong Yan

Decision tree-based probability estimation has received great attention because accurate probability estimation can possibly improve classification accuracy and probability-based ranking. In this paper, we aim to improve probability-based ranking under decision tree paradigms using AUC as the evaluation metric. We deploy a lazy probability estimator at each leaf to avoid uniform probability assignment. More importantly, the lazy probability estimator gives higher weights to the leaf samples closer to an unlabeled sample so that the probability estimation of this unlabeled sample is based on its similarities to those leaf samples. The motivation behind it is that ranking is a relative evaluation measurement among a set of samples, therefore, it is reasonable to yield the probability for an unlabeled sample with reference to its extent of similarities to its neighbors. The proposed new decision tree model, *LazyTree*, outperforms C4.5, its recent improvement C4.4 and their *state-of-the-art* variants in AUC on a large suite of benchmark sample sets.