Shrinkage on Decision Trees for Ranking

Bin Wang Huajie Zhang

An accurate ranking of instances based on their class probabilities is desired in many applications. In this paper, we propose to use shrinkage, a statistical technique, to improve the ranking performance of decision trees, measured by AUC (area under the Receiver Operating Characteristics curve). Shrinkage estimates the class probability of an instance using a linear interpolation of the local class probabilities on each node along the path from leaf to root. Thus, all nodes along the path are utilized, instead of only the leaf. We present a surprisingly efficient algorithm for learning the interpolating weights, which sets up the weights in one single iteration. In a traditional decision tree, one obstacle to yielding accurate ranking is that the instances falling into the same node are assigned the same class probability. We introduce a new method to generate distinct probability estimates by weighting instances differently. Our experiments show that various decision tree algorithms with shrinkage significantly outperform the original ones and other state-of-the-art techniques proposed to enhance the ranking performance of decision trees.