

ID3 is not an Occam algorithm

Balázs Szörényi

The general task of learning is to predict the outcomes of unseen events by processing some (more-or-less representative) examples. A well-known principle of learning theory is Occam's razor, which suggests, that if one can give a small representation or a short description for the examined activity of some kind of system (for example: customers' habits), then one can be quite sure, that this description can be successfully used for prediction. This principle is incorporated in the definition of the (α, β) -Occam algorithm.

The central question in computational learning theory is, whether a class \mathcal{C} of concepts ($c : X \rightarrow \{0, 1\}$ type functions for a given, common base set X) is PAC-learnable - that is: is there an L PAC-learning algorithm for \mathcal{C} ? It is known that any efficient (α, β) -Occam algorithm is an efficient PAC-learning algorithm, and there is also a certain converse of this result.

The PAC-learnability of decision trees is a major open problem. We study theoretical properties of the best-known learning algorithm in this special area, the ID3 algorithm. The result, we wish to present, is that ID3 is not an Occam-algorithm. This requires the construction of examples where ID3 constructs a much larger than optimal tree.