Rewarding misclassifications in oblique decision tree learning

András Salamon

Motivation to this work came from a real-world problem, the management of financial risks. Particularly, we had an interest in solving a credit granting (and monitoring) problem: the classification of loan applicants (and debtors).

Although credit granting process covers all the aspects of loan administration, we focus on only the classification process, when the funder institution (usually a bank) classifies the applicants. This classification should not only help to decide whether to give a loan or not, it should also help to determine the price of the loan, and should be included in the credit monitoring subprocess, where the classification is repeated — that time already of the loan debtors — on a regular basis, until the expiration of the loan.

We aimed at applying machine learning techniques for generating decision trees that are consistent with available data and domain knowledge. According to the Hungarian law, every bank should have a credit granting system, which has to be checked by the state authority. Because of the legal regulations, this system cannot be a black-box. Decision-tree algorithms can process the accumulated noisy historical data of the banks while using the predefined indicators of the legal regulations.

This kind of algorithms have been used since the 1960s. From the several variants of decision tree algorithms, we have chosen to use oblique decision trees. Unlike axis-parallel decision trees, oblique decision trees tests a linear combination of the attributes at each internal node. Although oblique decision tree learning requires more resource, we favor this, because it is more general than the axis-parallel variant.

We have enhanced oc1_v3, a well-known oblique decision tree algorithm. Oc1_v3 uses a top-down greedy tree building mechanism, based on the impurity measures of hyperplanes. For a given set of training examples, it cuts by introducing a hyperplane this set into two subsets so that the measure of the so-called impurity of the subsets be minimal. The actual impurity measure has a great impact on the structure and accuracy of the tree learned.

Although credit granting can be a binary classification problem, in our scenario we have five output classes. It is important to distinguish between different types of misclassifications. In our model, a so-called fitness matrix contains all the misclassification rewards: this matrix can be used in the accuracy calculation and in the tree building process.

Two impurity measures have been modified to use the fitness matrix. The new version of Sum Minority and Max Minority impurity measures uses the fitness matrix during hyperplane evaluation. Sum (Max) Minority measure calculates the sum (maximum) of the number of minority elements on both sides of the split. The new version counts the weighted number of minority elements, where the weight is based on the misclassification reward.

We made extensive experiments, working with both of the impurity measures on artificial — though in the credit granting domain realistic — datasets. The noise imposed on the training examples was systematically increased. In most of the cases the new tree learning methods, that rewarded misclassifications during the learning process, clearly dominated the original methods. A detailed confidence test showed that our results were significantly better in case of more noisy data. The test also showed that the enhancement has a greater impact on Max Minority compared to Sum Minority. On the other hand, the new methods were never surpassed significantly by the original tree learning methods.