

Rough Set Based Ensemble Classifier

C.A. Murthy, Suman Saha, and Sankar K. Pal

Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India
murthy@isical.ac.in

Combining the results of a number of individually trained classification systems to obtain a more accurate classifier is a widely used technique in pattern recognition. In [1], we introduced a Rough Set Meta classifier (RSM) to classify web pages. It tries to solve the problems of representing less redundant ensemble of classifiers and making reasonable decision from the predictions of ensemble classifiers, using rough set attribute reduction and rule generation methods on a granular meta data generated by base classifiers from input data.

The proposed method consists of two parts. In the first part, the outputs of individually trained classifiers are considered for constructing a decision table, with each instance corresponding to a single row. Predictions made by individual classifiers are used as condition attribute values and actual class – as decision attribute value. In the second part, rough set attribute reduction and rule generation processes are used on that decision table to construct a meta classifier. The combination of classifiers corresponding to the features of minimal reduct is taken to form classifier ensemble for RSM classifier system. Going further, from the obtained minimal reduct we compute decision rules by finding mapping between decision attribute and condition attributes. Decision rules obtained by rough set techniques are then applied to perform classification task.

It is shown that (1) the performance of the meta classifier is better than the performance of every constituent classifier, and (2) the meta classifier is optimal with respect to a quality measure that we proposed. Some other theoretical results on RSM and comparison with Bayes decision rule are also described. There are several ensemble classifiers available in literature like Adaboost, Bagging, Stacking. Experimental studies show that RSM improves accuracy of classification uniformly over some benchmark corpora and beats other ensemble approaches in accuracy by a decisive margin, thus demonstrating the theoretical results. Apart from this, it reduces the CPU load compared to other ensemble techniques by removing redundant classifiers from the combination.

References

1. Saha, S., Murthy, C.A., Pal, S.K.: Rough set based ensemble classifier for web page classification. *Fundamenta Informaticae* 76(1-2), 171–187 (2007)