3-level Confidence Voting Strategy for Dynamic Fusion-Selection of Classifier Ensembles

Csaba Főző and Csaba Gáspár-Papanek

Using *Classifier Ensembles* in data mining seems to be the easiest and most widespread way of further increasing classification accuracy nowdays. There are different stages of using multiple classifiers, such as the meta-classifier stage treating the ensemble members as black boxes, and the stage on which the functionality of members is also important. Furthermore, the method of combining ensemble members can be done by fusion or selection of classifiers. In this paper, we propose a novel procedure for constructing *Multiply Classifier Systems* (MCSs) [2] on the meta-classifier stage with an *Oracle* of a 3-level voting strategy. We use a dynamic, half fusion-half selection type method for ensemble member combination, which is a midway procedure between the extremes of fusion and selection. The MCS members are weighted and combined with the help of an Oracle. Classification is done by selecting the MCS's members that has weights higher than a predefined weight limit, and afterwards, fusing the remaining single classifier outputs. Due to this, the key role in our method is played by the Oracle, which is founded on a voting strategy of three levels: The explicit global (1), the implicit semi-global (2), and the local confidence (3). The first is calculated using the validation part of the labelled training data, and it is used to obtain the general accuracy of the single classifier. The second confidence segment is derived during the classifiers' construction via the implicit knowledge gathered simultaneous with training. Since this strongly depends on the internal operation of the classifier, it can not always be obtained, in example, when using some particularly complex classification methods. And finally, the third and most important part of the confidence triple is created with the help of the unlabelled object yet to be classified. Due to this, all of the MCS members have the global (1) and semi-global (2) confidence attributes, that can be counted in construction time, but the local one (3) must be derived in classification time. The procedure is similar to [1], with the difference that we are interested in the precise local confidences instead of the best single classifier. First, the k nearest neighbour of the unlabelled object (UO) is located by the help of an appropriate distance metric over data objects. Next, we select those, that has a Multiply Classifier Behaviour (MCB) similar to the UO. This is done by comparing the UO's vector with k decision vectors, whose elements are the decisions of MCS members for the given data object. Finally, the remaining n objects' MCBs are used to calculate the ratio of the correct decisions and n for each single classifier, resulting the local confidence. Combining the triplet of "general accuracy", the "self confidence", and the "expertise on the particular field" of the object yet to be labelled leads to a greatly increased accuracy in classification.

References

[1] Giorgio Giacinto, Fabio Roli. Dynamic classifier selection based on multiple classifier behaviour, *Pattern Recognition 34 (2001) 1879–1881*.

[2] Csaba Főző. State of the Art: Classifier Ensemble Methods.