



ADVANCES IN THE COMBINATION OF SUPERVISED CLASSIFICATION METHODS: AN EXPERIMENTAL STUDY

Sabina Mazza¹

ABSTRACT

In this work we were interested in investigating the predictive accuracy of one of the most popular learning schemes for the combination of supervised classification methods: the Stacking Technique proposed by Wolpert (1992) and consolidated by Ting and Witten, (1999) and Seewald (2002). In particular, we made reference to the StackingC (Seewald 2002) as a starting point for our analysis, to which some modifications and extensions were made. Since most of the research on ensembles of classifiers tends to demonstrate that this scheme can perform comparably to the best of the base classifiers as selected by cross-validation, if not better, this motivated us to investigate the performance of the Stacking empirically. An analysis of the results obtained by applying the our Stacking scheme, which includes differences and characteristic implementations compared to what is proposed by the literature, to the set of the dataset generated by means of an experimental design does not lead us to believe that the Stacking scheme is preferable in terms of performances to the use of the best single classifier. It always achieves good performances and is to be considered among the best. On the contrary, in the case of contaminated data, Stacking improves its performances noticeably, and generally appears to be very competitive, above all when the contaminations are more substantial.

Classification JEL: C 13, C 14, C 38

Keywords: Supervised classification methods, Ensemble learning, Stacking, Meta-level learning, Cross-validation

**Working Paper n° 133
October 2014**

¹ Sabina.mazza@uniroma1.it.

This paper is based on research work on “Combination of classification methods” supervised by Professor Giorgio Alleva.