DEPENDENCE OF CLASSIFIERS FUSION ACCURACY ON DIVERSITY OF CLASSIFIERS IN THE ANALYSES OF MEDICAL DATA

M. Cwiklinska -Jurkowska

Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland

† E-mail: *mjurkowska@cm.umk.pl*

In many applications of statistics in biomedical sciences the challenge is to classify new observations according to multiple correlated variables and classifiers built from these variables. However, the results may diverge substantially depending on the choice of specific classifiers. One option is to attempt a different procedure of the "fusion" of classifiers into one combined rule. Different character of constituent classifiers is usually the reason for the diversity of the ensemble. Diversity among individual classifiers of the ensemble is expected to be important for improving accuracy in classifiers fusion. Further, examining diversity measures might be useful for constructing the most efficient fixed combination of classifiers. Relationships between diversity and correctness of fixed combined classifiers, in simulated data, were examined by Kuncheva and Whitaker[2003]. We studied dependence of performance of fixed and also trained (with second-step level discrimination) combined classifiers on diversity, in real-life medical data.

We analysed big medical data sets. We examined classifiers' diversity by averaged pairwise measures and also by measures evaluating discrepancies between several classifiers.

To incorporate more diversity in the set of parametric and nonparametric Bayesian discriminant functions, we added also crisp classifiers, such as classification trees and neural networks. Different subsets containing constituent classifiers were examined. Spearman correlations between diversity measures and accuracy were stronger for symmetric diversity measures (when the diversity is the same if we swap values 1 and 0 for correct and incorrect decisions, respectively)] and we found higher dependence on diversity for fixed than for trained fusion.

Conclusion. Examining diversity might be useful in constructing efficient ensemble of classifiers.

References

L. I. Kuncheva, C. J. Whitaker Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. Machine Learning; 2003, 51 pp.181 - 207