Rose and TiMBL: a Match Made in Heaven

M. De Cock, E. E. Kerre

Department of Applied Mathematics and Computer Science

Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium

{Martine.DeCock, Etienne.Kerre}@rug.ac.be

In this paper we study the advantages of combining two software-packages:

The Tilburg Memory Based Learner (TiMBL) is developed at Tilburg University (the Netherlands) as a tool for (NLP)-classification purposes. The key-idea of memory based learning is the classification of new instances on the basis of their similarity to earlier instances which are stored in memory, rather than on the application of rules extracted from earlier instances.

The Rough Sets Data Explorer (Rose), created at the Poznan University of Technology (Poland), implements techniques based on the fundamentals of rough set theory. Among other things, this package allows for the selection of significant condition attributes (and hence reduction of redundant condition attributes) in an information table with classification examples (i.e. containing both condition and classification attributes).

To optimize the memory based learning process performed by TiMBL, we suggest to first scan a (small) table of training instances using Rose, looking for the significant attributes in the table. To complete and to store in memory a sufficiently large set of training examples, only these attributes have to be taken into account. Omitting the non-significant or redundant attributes is not only space and time saving for TiMBL, but can also lead to better classification results. In the paper we will illustrate this by means of the Dutch diminutive suffix problem that comes as an example with TiMBL.