Mutual Information for the selection of variables to model enantioselectivity

S. Caetano[1], C. Krier[2], M. Verleysen[2], Y. Vander Heyden[1]

[1]FABI, Department of Analytical Chemistry and Pharmaceutical Technology, Vrije Universiteit Brussel-VUB, Laarbeeklaan 103, 1090 Brussels, Belgium

[2] Université Catholique de Louvain, Machine Learning Group, DICE, 3 Place du Levant, 1348 Louvain-la-Neuve, Belgium

Molecular chirality is of fundamental interest in the pharmaceutical field where the identification and separation of enantiomers is of vital importance, since, in a biological medium, enantiomers should be regarded as different chemical compounds because they may show considerable differences in their interactions with other chiral molecules.

The mutual information (MI) criterion has been recently used to select relevant variables in spectrometric data [1, 2]. The MI measures the information content of **X** (explanatory variables) with respect to **y** (response variable), without making any assumption on the type of model that will be used.

This work will show some preliminary results of the application of MI to QSSR (Quantitative Structure-Selectivity Relationship) data. Therefore, from an initial set of molecular descriptors (≈ 1300), the MI criteria will select a much smaller set of variables (maximum 12), that will be then used to model the enantioselectivity of several molecules. Due to the fact that MI selects a rather small number of variables, the interpretability of the resulting model becomes easier.

[1] N. Benoudjit, D. François, M. Meurens, M. Verleysen, Chemometrics and Intelligent Laboratory Systems 74 (2004) 243-251

[2] F. Rossi, A. Landasse, D. François, V. Wertz, M. Verleysen, Chemometrics and Intelligent Laboratory Systems 80 (2006) 215-226