

Optimisation of the Performances of a Hybrid Expert System Designed for the Elucidation of Molecular Structures

S. Gosav, M. Praisler

*Department of Physics, "Dunărea de Jos" University, 111 Domneasca St, 800201 Galati, Romania
e-mail: sgosav@ugal.ro*

An expert system designed for the structure elucidation of volatile compounds having small molecular structures has been built by using a hybrid spectral database, formed by concatenating the GC-FTIR spectra with the GC-MS spectra of the molecules, in the attempt to benefit from both the selectivity of the infrared spectra and of the sensitivity of the mass spectra. On the other hand, characterizing each sample by such a hybrid spectrum has the disadvantage of decreasing the ratio between the number of samples and the number of input variables. In many cases, the number of samples included in the training, test, or validation set cannot be increased, simply because the number of compounds forming the class of substances to be modeled is limited to a relatively small number of analogues. In this case, the challenge in enhancing the performances of the expert system by optimizing the sample/variable ratio is to find efficient selection criteria, which would allow the reduction of the number of input variables without losing the relevant information. In other words, these criteria should allow us to eliminate *only* the redundant information and to maintain into the system (as much as possible) the spectral information needed to obtain a good modeling power.

In this paper we are analyzing two such criteria, i.e. the importance and the sensitivity of the input variables. The advantages and disadvantages of each of these two criteria are also analyzed by comparing the performances of the optimized expert system with those obtained when applied for expert systems built with homogenous spectral data-bases (containing only GC-FTIR or only GC-MS data).