Supervised Self-Organising Maps

Ron Wehrens, Egon Willighagen,
Willem Melssen and Lutgarde Buydens
Radboud University Nijmegen,
The Netherlands

Self-organising maps (SOMs) provide a means to project a collection of possibly high-dimensional data points to a two-dimensional grid, preserving topology. This means that objects mapped in the same region are similar. SOMs have been used for many different applications, mapping anything from a hundred to hundreds of millions of objects. Typically, the mapping is done in an unsupervised fashion. An example from chemistry is the analysis of the Cambridge Structural Database, which contains almost 400,000 crystal structures. Using a specially designed similarity function it is possible to project these (or a relevant subset) to a SOM. This mapping can be used in several different ways: it can help in identifying the crystal structure of unknown compounds using experimental data, it provides a visual and appealing way of inspecting the database, it has applications in quality control, etcetera.

In this example, the crystal structures are represented by X-ray powder patterns. In the natural sciences one typically has a situation where additional information, either class information or additional continuous variables, is available. In the example of the crystal structures, extra variables are the symmetry class of the crystal structure, the molecular volume, and others. We show how to combine these in a straightforward and consistent way, and present two new forms of the self-organising map that can efficiently incorporate this information in the mapping. A specific advantage of this kind of supervised mapping is the interpretability of the results.