VISIPOINT

Software/SOM and Sammon’s Mapping Technologies

One of the best known unsupervised neural learning algorithms is the Self-Organizing Map (SOM) (Kohonen, 1995), the aim of which is to find prototype vectors that can represent the input data set and at the same time to achieve a continuous mapping from the input space to a lattice. The computing time of the original SOM increases linearly when the number of reference vectors (neurons) is increased.

Therefore, our software uses fast and easy to use Tree-Structured Self-Organizing Map (TS-SOM), a variant of the SOM (Koikkalainen, 1994, Häkkinen, 2001). The computing time of TS-SOM increases only logarithmically and in some applications might even be almost independent of the number of neurons of the network.

The weight vectors of the SOM are first initialised to random values. With each training pattern the winning neuron (Best-Matching Unit, BMU) is first found by comparing the input (measured) and weight vectors of the neurons by Euclidean distance metrics. The weights of the winning neuron (BMU) and its four neighbours (in the TS-SOM implementation) are then moved towards the input vector according to a learning rate factor which decreases monotonically towards the end of learning. The resulting map (consisting of weight vectors) can be visualized and it can also be used as starting point of further analysis like Sammon's mapping or fuzzy description of phenomena found from the measured data.

Sammon's mapping is an iterative method based on a gradient search (Sammon Jr, 1969). The aim is to map points in n-dimensional space usually into 2 dimensions. The algorithm finds the locations in the target space so that as much as possible of the original structure of the measurement vectors in the n-dimensional space is conserved. The numerical calculation is more time-consuming than the SOM algorithm, however, which can be a problem with a massive data set. On the other hand, it is able to represent the relative distances between vectors in a measurement space and is thus useful for determining the shape of clusters and the relative distances between them. It is therefore of benefit to combine these two algorithms. Sammon's mapping is thus applied to the stage where the TS-SOM algorithm has already achieved a substantial data reduction by replacing the original data vectors with a smaller number of representative prototype vectors

References:

Häkkinen E., Design, Implementation and Evaluation of Neural Data Analysis Environment. Jyväskylä Studies in Computing 12, University of Jyväskylä, 2001, PhD Thesis.

Kohonen T., Self-Organizing Maps, Springer-Verlag, Berlin, 1995.

Koikkalainen P., Progress with the tree-structured self-organizing map. Proceedings of ECAI'94, 11th European Conference on Artificial Intelligence (Cohn, A. G., Ed.), Wiley and Sons, New York. 1994, pp. 211-215.

Sammon Jr. J.W, A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.* **18**,1969, pp. 401–409.