

Self-organizing map browser for database retrieval

*Antti Kerminen, Antti Raike, Mauri Kaipainen**

University of Art and Design Helsinki, Media Lab, Soft Computing Interfaces Group
mauri.kaipainen@uiah.fi

We introduce a web tool to access database information by clicking and dragging on similarity-clustered item labels on a two-dimensional map visualization of a multi-dimensional data corpus, formed using the self-organizing map (SOM) algorithm.

Usability problem with one-dimensionally displayed information

Traditional presentations of data, such as retrieved from databases or internet typically force the user to view data ordered by one dimension, such as date of creation or alphabets, or as hierarchical embedding of items, such as the system of folders on computer desktop. From the users' viewpoint, such presentations require a level of academic skills to be grasped, due to poor match with user's cognitive map. This may be because they call for attention on one dimension at a time. Such presentations fail to optimize viewing information in which significant patterns correspond to holistic matches satisfying multiple criteria.

As an example from human life, for most people in search for the only one it is not sufficient to rank the candidates solely by beauty, or by richness. Instead, the natural way is to search for holistic co-occurrence of beauty, richness, wisdom and generosity, beside other desired virtues. Other examples of such match include surveys that result in collections of qualitative judgments of relevance or importance of given properties, or degrees of category membership, as well as hard technological control tasks with multiple indicators of system-state to be simultaneously monitored, such as airplane cockpits.

The cognitive load in such multi-criteria match tasks can be reduced by presenting the user with results that are pre-clustered by multiple criteria. From the point of view of matching users' cognitive capabilities, it may not be irrelevant that perceptual or behavioral patterns are known to be responded to by fields of neural activity on the cortex, with a point of maximum and graded degrees of activity distributed widely around them. Relatively similar activity fields correspond to relatively similar patterns. Tonotopies (Hood 1977), somatotopies (Merzenich 1988; Wall 1988) and spatial representations (Olton 1977) are examples of such cortical mappings.

Self-organizing map

The output of the self-organizing map (SOM) introduced by Kohonen (1984), is reminiscent of such cortical responsivity. It is an artificial neural network based on very simple neural-like assumptions. It offers a method of ordering and visualizing complex data by similarity, based on multiple criteria. From the point of view of statistics, it is conceivable as a model of how the brain may implement multidimensional scaling. The SOM relies solely on unsupervised learning. As input, it accepts data described as vectors of feature strengths. On the resulting map there is a localized focus for each data item. Furthermore, the map can be tested for any configuration of properties.

Description of the interface

The SOM-based interface developed by the authors lets the user retrieve information in a manner native to human cognition, as we claim. The interface, implemented as JAVA

applet, consists of the map, the property panel and the data panel. In addition, there is a separate display window for the data to be retrieved.

By default, *the map panel* displays the labels of data items, each located at the point of its maximum response. The background coloring corresponds to the degree to which chosen properties represented by each map location. The user can drag over the items or an area in order to view the data corresponding to the items located within it. *The property panel* lets the user choose the properties of whose combined response field is to be plotted on the map with graded degrees of a dedicated color. *The display panel* lets the user choose the information associated with the items chosen in the map panel.

Example of use

As an example of potential applications we use a thematic profile data derived from a doctoral study database. The data consists of doctoral students' judgments of the relevance of themes like "ethical concerns", or "human-machine interaction" to their doctoral projects. The judgments assume one of values "not relevant", "not very-relevant", "fairly relevant" and "relevant", corresponding to values 0, .33, .67, and 1, respectively. A self-organizing map was computed using the data and the standard procedure, as described by Kohonen (1984).

Clicking at the options given in the property panel, it is easy to view the projects to which the chosen theme or combination of themata is relevant. By dragging over the map, the user can choose a subset of the projects to be examined. The information to be shown in the display panel, e.g. contact information or abstract, can be chosen in the data panel.

Conclusion

The introduced prototype of the SOM-based browser takes the advantage of how the self-organized map arranges multi-dimensionally varying data into similarity clusters, with the intention to facilitate information retrieval from such data in which significance of hits is defined by multiple properties as a whole.

Literature

- Hood, J. (1977). *Psychological and Psychological Aspects of Hearing*, -Critchley, M.; Henson, R. 1980. Music and the brain. London: Heinemannss.
- Kohonen, T. (1984). *Self-Organisation and Associative Memory*, Berlin etc.: Springer-Verlag.
- Merzenich, M. M.; Recanzone, G.; Jenkins, W. M.; Allard, T. T.; Nudo, R. J. (1988). *Cortical Representational Plasticity*, -Rakic, P.; Singer, W. (Eds.) 1988. Neurobiology of Neocortex. John Wiley & Sons Limited.
- Wall, J. T. (1988). Variable organization in cortical maps of the skin as an indication of the lifelong adaptive capabilities of circuits in the mammalian brain, Trends in Neuroscience, Vol. 11, No. 12.

Acknowledgement

This study was funded by the National Technology Agency of Finland (TEKES decision 40691/00).