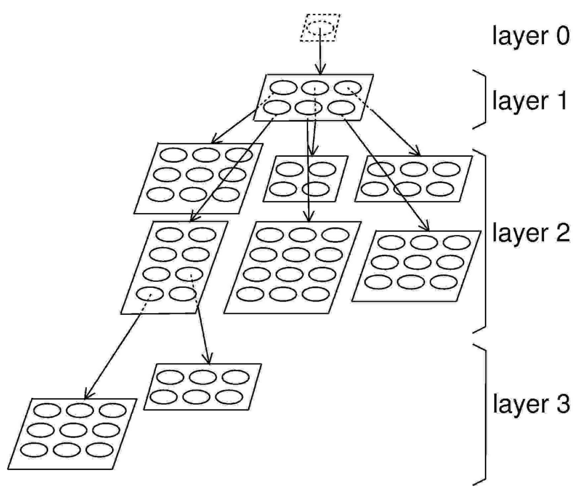
Growing Hierarchical Self Organizing Maps (GHSOM)

Self Organizing Maps (SOM's) are great tools for neural network models used in many data mining applications. They map high dimensional data so that similar data are located close to each other using unsupervised methods. Although SOM's computation efficiency and intuitive way of present data is very useful, there are a few limitations with them. The first is that the SOM uses a fixed network architecture. This is a problem when dealing with large amounts of data because the number and arrangements of elements must be predefined prior to training the SOM. Another issue is that SOM's lack the ability to extend the hierarchical structure of data. If it was extended, more detail would be given especially if certain clusters had a high density of data. Both of these limitations are solved by using Growing Hierarchical Self Organizing Maps (GHSOM's).



While the SOM has great capabilities, the GHSOM enhances these by using an incremental growing method for the maps size and being able to adapt to the hierarchical structures in the data. The key idea behind GHSOM is that it uses a multilayer hierarchical structure consisting of independent growing SOM's. By dynamically fitting the multilayered architecture according to the structure of the data, the issue of architecture not hierarchically adaptive and the fixed size of the map.

The major benefits of GHSOM model compared with the standard SOM are the following. First, the overall training time is largely reduced since only the necessary number of units are developed to organize the data collection at a certain degree of detail. Second, the GHSOM uncovers the hierarchical structure of the data by its very architecture, thus allowing the user to understand and analyze large amounts of data in an explorative way. Third, with the various emerging maps at each layer of the hierarchy being rather small in size, it is much easier for the user to keep an overview of the various clusters. Last, but not least, by ensuring a consistent global orientation of the individual maps in the respective layers, the topological similarities of neighboring maps are preserved. Thus, navigation across map boundaries is facilitated, allowing the exploration of similar clusters that are represented by neighboring branches in the GHSOM structure.

It is clear that GHSOM solves many of the issues with SOM's, mainly the size of the map. Aside from a having a dynamic map size, the overall train time is reduced. With the addition of TMR, less tests are need to be ran because the parameter value is set automatically as well it improves decision making on where to insert units. One draw back is that TMR uses significantly more resources, but because it needs to run less tests, it is still more efficient. While most experiments use a small amount of variables, it would be worth it try this method with a larger amount.

References

[1] The Growing Hierarchical Self-Organizing Map http://www.ifs.tuwien.ac.at/~andi/ghsom/index.html

[2] A New Approach to Hierarchical Clustering and Structuring of Data with Self-Organizing Maps http://rose.ofai.at/~elias.pampalk/publications/pam_ida04.pdf

[3] The Growing Hierarchical Self-Organizing Map: Exploratory Analysis of High-Dimensional Data http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1058070