

Abstract

Self-organizing maps (SOM) are self-organized projections of high dimensional data onto a two dimensional map wherein vector similarity is implicitly translated into topological closeness in 2D projection. They are thus used for visualization and clustering of high dimensional data. However it is often challenging to interpret the results due to drawbacks of currently used methods for identifying and visualizing clusters in the resulting feature map. In this thesis we introduce a new phase to the SOM that advances the cluster separation process by strengthening cluster boundaries, which we refer to as Cluster Reinforcement (CR) phase. The CR-phase helps the weights of the same clusters become more similar; consequently cluster boundaries are much more evident. We also define a new Boundary (B) matrix method for visualization of clusters in feature maps. The B-matrix method not only makes cluster boundaries sharp and clear, but also can be overlaid on maps of component planes (something that was not possible with previous methods). The combination of the SOM, CR-phase and B-matrix results in highly informative and user-friendly identification and visualization of clusters in high dimensional data. We demonstrate this methods on three data sets: classic animal benchmark test, an actual phonetic word clustering problem, and clustering of geographic data related to fuel efficiency of vehicle choice.