

Abstract

Extending the Growing Hierarchical SOM for Clustering Documents in Graphs domain

The growing hierarchical self-organizing map (ghsom) is the most efficient model among the variants of som. It is used successfully in document clustering in various pattern recognition applications effectively. The main constraint that limits the implementation of this model in all the other variants of som models is that they work only with vector space model (vsm). In this paper, we extend the ghsom to work in the graph domain to enhance the quality of clusters. Specifically, we represent the documents by graphs then cluster those documents by using a new algorithm g-ghsom: graph-based growing hierarchical som after modifying its operations to work with the graph instead of vector space. We have tested the g-ghsom on two different document collections using three different measures for evaluating clustering quality. The experimental results of the proposed g-ghsom show an improvement in terms of clustering quality compared to classical ghsom.