Sequential hierarchical pattern clustering

Farran, B. , Ramanan, A. and Niranjan, M.

School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom

Abstract

Clustering is a widely used unsupervised data analysis technique in machine learning. However, a common requirement amongst many existing clustering methods is that all pairwise distances between patterns must be computed in advance. This makes it computationally expensive and difficult to cope with large scale data used in several applications, such as in bioinformatics. In this paper we propose a novel sequential hierarchical clustering technique that initially builds a hierarchical tree from a small fraction of the entire data, while the remaining data is processed sequentially and the tree adapted constructively. Preliminary results using this approach show that the quality of the clusters obtained does not degrade while reducing the computational needs.

Author keywords

Gene expression; Hierarchical clustering; Large scale data; On-line clustering

Indexed keywords

Clustering methods; Hierarchical clustering; Hierarchical patterns; Hierarchical tree; Large scale data; Machine-learning; On-line clustering; Pairwise distances; Unsupervised data

Engineering controlled terms: Bioactivity; Bioinformatics; Gene expression; Pattern recognition

Engineering main heading: Cluster analysis