Referential hierarchical clustering algorithm based upon principal component analysis and genetic algorithm

Jui-Shin Lin ∗
Department of Technology Management and Engineering Overseas
Chinese Institute of Technology
Taichung, Taiwan
R.O.C.

Tung-Shou Chen
Yung-Hung Kao
Chih-Chiang Lin
Yung-Hsing Chiu
Graduate School of Computer Science and Information Technology
National Taichung Institute of Technology
Taichung, Taiwan
R.O.C.

Abstract

The Hierarchical Clustering (HC) is not designed to locate the leaf nodes in the tree structure, and therefore is not suitable to locate similarity relation on the sequence of the leaf nodes. In order to generate the similarity relation on tree structure diagram of HC, we proposed an improved solution; the Referential Hierarchical clustering Algorithm (RHA). RHA is a combination of HC, Genetic Algorithm (GA) and Principal Component Analysis (PCA) to resolve the problem of the traditional HC. PCA is a technique that reduces high-dimensional dataset to lower dimensions for analysis and reconstructs each data by a suitable linear combination of the principal components. These principal components are ordered by the amount of the variance which is explained in the original dataset. Therefore, RHA adopts GA to find the solution which has the same tree structure with HC and the most similar with the sequence of the samples sorted by increasing value of the first principle components. Experimental results show that the clustering result of RHA exposes the similarity relations between the leaf nodes and the clusters. RHA could be applied to any problem in HC and the generated tree diagram could assist researchers to compare and analyze each sample and find the relations between the clusters more easily and quickly.

Keywords and phrases: Hierarchical Clustering (HC), Principal Component Analysis (PCA), Genetic Algorithm (GA), Tree Structure Diagram, Similarity Relation.

∗E-mail: jslin@ocit.edu.tw

Journal of Statistics & Management Systems
Vol. 13 (2010), No. 2, pp. 255–265
© Taru Publications