Hierarchical Clustering for Mixed Feature-Type Complex Data

Huiwen Wang
School of Economics and Management, Beihang University, Beijing 100191, China - wanghw@vip.sina.com
Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beijing 100191, China

Cheng Wang*
School of Economics and Management, Beihang University, Beijing 100191, China - wangcheng@buaa.edu.cn
Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beijing 100191, China

Yuan Wei
School of Economics and Management, Beihang University, Beijing 100191, China - weiyuan1143@163.com
Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beijing 100191, China

With the rapid development of cross-platform data collection technology and the coming of the big data era, there are always a mixture of single-valued data, symbolic data, composition data and functional data in one table, which can be called mixed feature-type complex data. So far no operating rules among these different types of data have been given, thus no multivariate models based on mixed data can be built, which has already become a major obstacle of data analysis in economics and management. In many real applications that involve mixed feature-type complex data, clustering is needed to classify the information. In this paper, we present dissimilarity matrix on the multidimensional mixed feature-type complex space to measure the distance between different observations, and propose agglomerative hierarchical clustering method for mixed feature-type complex data based on the dissimilarity matrix. Experiments with simulated data in a framework of a Monte Carlo schema are proposed to verify the performance of the hierarchical clustering method based on the dissimilarity matrix.

Keywords: Dissimilarity Matrix; Symbolic Data; Composition Data; Functional Data.