

Exploratory data analysis of interval-valued symbolic data with matrix visualization

Chiun-How Kao^{1,2}, Junji Nakano³, Sheau-Hue Shieh⁴,
Yin-Jing Tien^{2*} (田銀錦), Han-Ming Wu⁵, Chuan-kai Yang¹
and Chun-houh Chen²

¹Department of Information Management, National Taiwan University of Science and Technology, Taiwan

²Institute of Statistical Science, Academia Sinica, Taiwan

³Department of Data Science, The Institute of Statistical Mathematics, Japan

⁴Center for Educational Program for Secondary School Teachers, National Taipei University, Taiwan

⁵Department of Mathematics, Tamkang University, Taiwan

Abstract

Symbolic data analysis (SDA) has gained popularity over the past few years because of its potential for handling data having a dependent and hierarchical nature. Amongst many methods for analyzing SDA data, exploratory data analysis (EDA: Tukey, (1977)) with graphical presentation is an important one. Recent developments of graphical and visualization tools for SDA include zoom star, closed shapes, and parallel-coordinate-plots. Other studies project high dimensional SDA data into lower dimensional space using SDA versions of principal component analysis, multidimensional scaling, and self-organizing maps. Most graphical and visualization approaches for exploring SDA data structure inherit the advantages of their counterparts for conventional (non-SDA) data, but also their disadvantages. Here we introduce matrix visualization (MV) for visualizing and clustering SDA data using interval-valued symbolic data as an example; it is by far the most popular SDA data type in the literature and the most commonly encountered one in practice. Many MV techniques for visualizing and clustering conventional data are converted to SDA data, and several techniques are newly developed for SDA data. Various examples of data with simple to complex structures are brought in to illustrate the proposed methods.

Keywords: symbolic data analysis, interval-valued data, matrix visualization, generalized association plots, proximity matrix, exploratory data analysis, EDA