An Extension of Parallel Coordinate Plot for Visualizing Aggregated Symbolic Data

Yoshikazu YAMAMOTO

Tokushima Bunri University, Kagawa, Japan yamamoto@fe.bunri-u.ac.jp

Junji NAKANO

The Institute of Statistical Mathematics, Tokyo, Japan nakanoj@ism.ac.jp

Symbolic data is a second level data to express a concept or a group of individuals, and is typically described by intervals, histograms or barcharts as values of variables. Note that these values indicate information about the marginal distribution of each variable. We consider the case where all individuals in a group are available and we can calculate any statistics which express information about the joint distribution of variables. We call such summarization of a group as aggregated symbolic data.

We consider the visualization of aggregated symbolic data described by means, standard deviations and correlation coefficients of variables. We used an extended parallel coordinate plot for visualizing these statistics of groups at the same time. Simply visualizing all correlation coefficients is, however, sometimes complicated and difficult to interpret. In order to overcome this difficulty, we propose using principal components of variables in a group for visualizing aggregated symbolic data. We have developed an interactive statistical graphics software that has interactive operations such as selection and linked highlighting to visualize aggregated symbolic data intuitively.

Key Words: Data reduction, Interactive statistical graphics, Parallel coordinate plot, Principal component analysis