

Davis-Jasp: a data mining solution by combining two separate Java-based systems

Huh, Moon Yul

Department of Statistics, Sungkyunkwan University, Seoul, Korea
myhuh@skku.ac.kr

Song, Kwang Ryeol

Department of Statistics, Sungkyunkwan University, Seoul, Korea
skr9697@korea.com

Yamamoto, Yoshikazu *Tokushima Bunri University, Kagawa, Japan*

yamamoto@is.bunri-u.ac.jp

Nakano, Junji *The Institute of Statistical Mathematics*

4-6-7 Minami-azabu, Minato-ku, Tokyo 106-8569, Japan
nakanoj@ism.ac.jp

Fujiwara, Takeshi *The Graduate University for Advanced Studies*

4-6-7 Minami-azabu, Minato-ku, Tokyo 106-8569, Japan
fuji@ism.ac.jp

Kobayashi, Ikunori *Faculty of Engineering, Tokushima Bunri University*

Shido 1314-1, Okawa, Kagawa 769-2193, Japan
ikunori@es.bunri-u.ac.jp

1. Introduction

Visualizing data is one of the common methods in data mining techniques. Information contained in huge data set sometimes are visible by appropriate graphics, because we are accustomed to grasp patterns from visual displays.

Data visualization techniques have been mainly developed in the statistics community as descriptive statistical methods from the beginning of that history. Bar charts, histograms or scatterplots have been used as useful graphics. In the exploratory data analysis techniques, some new graphs, for example, stem and leaf plots or boxplots, were proposed and have been widely used. Recently, with the development of computer technologies, these graphs have been sophisticated and improved much by using computer power and its advanced graphical user interface, for example, scatterplot matrix or changing the number of bins of histogram

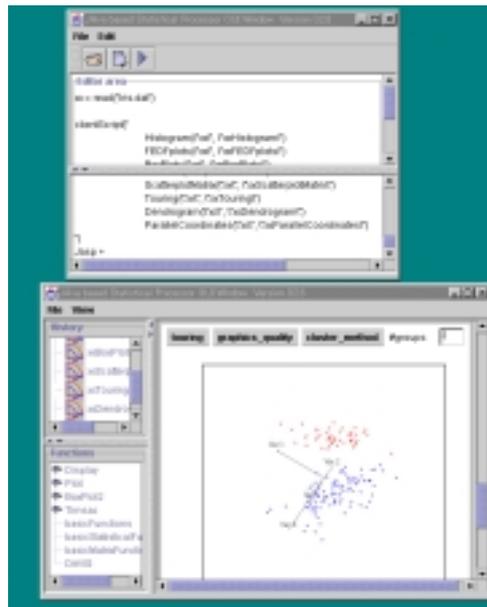


Figure 1: Executing Davis functions in Jasp: CUI and GUI windows of Jasp user interface are shown. A program for using Davis functions are called on the CUI editor window and was executed. Results are given on the GUI window, which shows Grand Tour here.

dynamically. New methods, which utilize new computer technologies fully and extensively, such as 3D dynamic graphics, have also been developed.

Davis (Data visualization system) was developed by Huh and Song (<http://stat.skku.ac.kr/~myhuh/software/DAVIS/DAVIS.htm>). This has been implemented to handle various graphical methods for statistical analysis and data mining works. Davis is written in Java language. Because of the mechanism of Java language execution, it is platform independent, i.e., it runs on many operating systems such as Windows and UNIX (including Linux). Davis is a stand-alone application and is designed for easy operation by graphical user interface (GUI).

Jasp (Java-based statistical processor) was separately developed by Nakano, et al. (2000) for a general purpose statistical and data mining system. Jasp is also written in Java. Jasp has many advanced facilities, such as function based, object oriented language, closely connected character user interface and graphical user interface, distributed computing abilities and foreign language interface. As Jasp and Davis are all Java-based systems, and their objectives are complementary, we thought it would be nice to integrate these two separate systems to develop into one integrated system. In this talk, we present how we realized this integration, and will show the capabilities of this new system. In this way we made it possible to use tools of Davis from within Jasp so that our new system could be a practical data mining tool that can handle not only the tools for data filtering and statistical analysis but also the tools for data visualization.