Journal of Universal Computer Science, vol. 11, no. 11 (2005), 1749-1751
submitted: 17/11/05, accepted: 22/11/05, appeared: 28/11/05 © *J.UCS*

Visual Data Mining
J.UCS Special Issue

Jesús S. Aguilar–Ruiz
(Polytechnic, Pablo de Olavide University, Spain
direscinf@upo.es)

Francisco J. Ferrer–Troyano
(Computer Science Department, University of Seville, Spain
ferrer@lsi.us.es)

Visualization has proven to be an essential support throughout the KDD process in order to extract hidden information from huge amount of data. Visualization techniques enable the direct integration of the user to overcome major problems of automatic machine learning algorithms methods such as presentation and interpretation of results, lack of acceptance of the discovered findings, or limited confidence in these. Since computers are still much less useful than the ability of the human eye for pattern matching, visual data exploration techniques provide the user with graphic views or metaphors that represent potential patterns and data relationships. These techniques allow to combine human visual perception and recognition capabilities with the computational power of today's computer systems, making easier to detect interesting patterns and trends in the data. Human perception can straightforwardly identify data relationships when a data set is two or three–dimensional. However, multidimensional data sets with more than three dimensions require some kind of visual transformation to be explored and analysed. In addition, the ever–increasing input data size leads to new challenges on visualization techniques and concepts. Due to technological progress, scientific and commercial applications are capable of generating, storing, and processing massive amounts of data. Therefore, new visualization techniques ought to scale well on very large data sets and avoid visual metaphors suffer from the data size. In this special issue, the most recent Visual Data Mining approaches to such problems are included according to the next order.

Martin Atzmueller and Frank Puppe present a novel semi–automatic approach for visual and explorative subgroup mining implemented in the VIKAMINE system. They discuss and describe how the subgroup mining process can be improved by user integration, and propose utilizing the zoomtable technique as the key visualization method to guide such a process.

Daniel A. Keim and Jörn Schneidewind introduce a new approach based on a Multiresolution paradigm to increase the scalability of existing Visual Data

Exploration techniques. This approach is based on a given relevance function in order to present data at different granularities. From this function, highly relevant objects can be presented at full detail and less relevant objects at lower detail. The available display space is then distributed according to the data granularity, to emphasize relevant information.

Jason J. Jung claims a collaborative filtering system based on information propagation in distributed social network environment and proposes a novel method for visual explanation of the recommender system on social network. In contrast with centralized recommender systems, social recommendation algorithm is applied to the item rating data on social networks. Meaningful recommendation can be uncovered by the topology of social network as well as the similarity between users, becoming propagated into the users in the estimated same groups.

Klaus Hinum et al. present a novel spring-based interactive visualization method for analysing data derived from questionnaires involves a number of highly structured, time-oriented parameters. In addition, the authors address the particular problem with cognitive behavioral treatment (CBT) of anorexia nervosa in adolescent girls.

César García-Osario and Colin Fyfe present and extension of Andrews' curves by a moving three-dimensional image in which the user can see clouds of data points moving as he move along the curves. The combined use of this new display with techniques such as brushing and linking allows the user to identify visually the clusters and outliers present in high-dimensional data.

Li Wei et al. introduce a novel framework that allows visualization to take place in the background of day-to-day computer use. This system works by replacing the standard file icons with automatically generated Intelligent Icons that reflect the contents of the files in a principled way. While there is little utility in examining an individual icon, examining groups of them provides a greater possibility of unexpected and serendipitous discoveries. The authors demonstrate the utility of their approach on data as diverse as DNA, text files, electrocardiograms, and Space Shuttle telemetry.

Francisco J. Ferrer et al. describe a interactive visual exploration technique approaching two data mining task generally addressed by independent machine learning algorithms: classification and feature selection. Through different metaphors with dynamic properties, this technique allows the user can re-explore meaningful intervals belonging to the most relevant attributes, interactively building decision rules and increasing the model accuracy through successive exploration levels.

Finally, Denis V. Popel remarks the size explosion problem through defining uncertain values as the major obstacle in the representation and visualization of incompletely specified data, which commonly raises questions about suggested heuristics and their practical applicability. The author emphasizes the renewed interest in resolving this problem based on symbolic techniques and gives an outline of decision diagrams for representing incomplete and uncertain dependencies.

Jesús S. Aguilar–Ruiz
Francisco J. Ferrer–Troyano
(Sevilla, November 2005)