Visual Data Mining – KDD Workshop Report

Stephen G. Eick
Visual Insights
Naperville, IL 60563

eick@visualinsights.com

Daniel A. Keim
AT&T Research and Constance University
Florham Park, NJ 07932

keim@att.research.com

On August 26, 2001 a one-day workshop on Visual Data Mining was held in conjunction with the KDD-2001 conference. About 50 people attended the workshop, mostly from industry with some academics. The audience included both data mining algorithm experts and visualization specialists. During the workshop thirteen peer-reviewed papers were presented that treated both problem-specific issues as well as broader topics.

The focus of the workshop was to explore how to include humans in the data mining process. The idea is that by combining human flexibility, creativity, and general knowledge with the powerful computational capability of today's computer systems, we can increase the effectiveness of both. This approach, known as Visual Data Mining, has proven to be particularly effective for exploratory data analysis when the goals of the analysis are vague. Visualization enhances automated KDD algorithms by increasing user involvement and thereby user confidence in the findings. This makes Visual Data mining an indispensable component of the data mining process.

Some of the highlights from the technical agenda included:

- High dimensional visualization – survey papers by Grinstein (High-dimensional Visualization) and Inselberg (Data Mining and Visualization of High Dimensions)

- Pixel-oriented techniques – Ankerst (Visual Data Mining with Pixel-oriented Visualization Techniques) and Hao, et al. (Visual Mining of E-Customer Behavior using Pixel Bar Charts)

- Clustering and classification – Davison, et al. (A Particle Visualization Framework for Clustering and Anomaly Detection), Caragea, et al. (Gaining Insights into Support Vector Machine Classifiers), and Strehl, et al. (Relationship-based Visualization of High-dimensional Data Clusters)

- Visual Frameworks – Simoff (Form-Semantics-Function – A Framework for Designing Visualization

Models for Visual Data Mining) and Kopanakis, et al. (Visual Data Mining & Modeling Techniques)

- Combining Visualization and Data Mining – Zho, et al. (Visual Analysis of the Behavior of Discovered Rules)

- Domain-specific applications – Deleus, et al. (Science and technology interactions discovered with a new topographic map-based visualization tool), Mazeika, et al. (Surface Computations for Immersive Explorative Data Analyses), Ma, et al. (EventMiner: An integrated mining tool for scalable analyses of Event data)

Three crosscutting themes emerged from the rich discussions that surrounded the paper sessions and emerged during the breaks. The first involved scalability and how to handle large datasets. The approaches considered addressed dimensional scalability, reducing data sizes via algorithms, and packing more information into the visual presentation by exploiting pixel-based techniques. The second involved how best to **combine visualization and data mining** algorithms. The third focused on the highly useful examples and how to obtain **general research results from the domain-specific applications.**

From the lively discussions it was clear that the workshop participants were highly enthusiastic and excited about the topics considered. It was generally concluded that combining visualization and data mining will increase the effectiveness and impact of both.

The full workshop program can be found at

http://www.inf.uni-konstanz.de/~keim/KDD_Workshop/KDDWorkshopProgram.html

and the workshop notes are available via

http://www.inf.uni-konstanz.de/~keim/KDD_Workshop/KDD_Proceeding.pdf.