International Journal of Information Movement Vol.2 Issue IX January 2018)

Website: www.ijim.in ISSN: 2456-0553 (online) Pages 12-14

DATA MINING – A VISUAL TECHNIQUE

Shivali Gupta
M.Lib.I.Sc. (University of Jammu), NET, JKSET
Librarian (Govt. College for Women, Udhampur (J&K)
E-mail: shivali.jandial1990@gmail.com

Arun Gupta
M.Lib.I.Sc. (University of Jammu), NET
Chief Librarian (Govt. Degree College, Udhampur (J&K)
E-mail: arun_gupta12376@yahoo.in

Abstract: This paper describes a number of important visual data mining issues and introduces techniques employed to improve the understandability of the results of data mining. Firstly, the visualization of data prior to, and during, data mining is addressed. Through data visualization, the quality of data can be assessed throughout the knowledge discovery process. Secondly, information visualization is discussed, i.e. how the knowledge, as discovered by a data mining tool, may be visualized throughout the data mining process. In addition, the paper shows how virtual reality and collaborative virtual environments may be used to obtain an immersive perspective of the data and the data mining process.

Keywords: Data Mining, Data Visualization, Information Visualization, Virtual Reality, Visual Data Mining

1.0 Introduction : The current explosion of data and information, which are mainly caused by the continuous adoption of data warehouses and the extensive use of the Internet and its related technologies, has increased the urgent need for the development of the techniques for intelligent data analysis. Data mining automates the discovery of hidden patterns and relationships that may not always be obvious. Data mining tools include classification techniques, clustering algorithms and association rule approaches etc. Data mining has been fruitfully used in many of domains include marketing, medicine, finance, engineering, and bioinformatics. Due to which, the result of many data mining techniques are often difficult to understand. For example, the results of a data mining effort producing 300 pages of rules will be difficult to analyze. The visual representation of the knowledge embedded in such rules will help to heighten the comprehensibility of the results. The visualization of the data itself, as well as the data mining process should go a long way towards increasing the user's understanding of and faith in the data mining process.

1.1 History

Human beings intuitively search for novel features, patterns, trends, outliers and relationship in data. Through visualizing the data and the concept descriptions obtained, a qualitative overview of large and complex data sets can be obtained. In addition, data and rule visualization can assist in identifying regions of interest and appropriate parameters for more focused quantitative analysis. The use of data and rule visualization thus greatly expands the range of models that can be understood by the user.

Data mining techniques construct a model of the data through repetitive calculation to find statistically significant relationships within the data. However, the human perception system can detect patterns within the data that are unknown to a data mining tool. This combination make strengths of the human visual system and data mining tools may lead to the improvement of human's perspective of the problem at hand.

Visual data mining integrates data visualization and data mining and is closely related to computer interfaces, pattern recognition and high performance computing.

2.0 Data and Information Visualization

2.1 Data Visualization Data visualization provides a powerful mechanism to aid the user during both data preprocessing and the actual data mining. For example, large samples can be visualized and analyzed. In particular, visualization may be used for outlier detection, which highlights surprises in the data. In addition, the user is aided in selecting the appropriate data through a visual interface. Data transformation is an important data preprocessing step. During data transformation, visualizing the data can help the user to ensure the correctness of the transformation. Visualizing may also be used to assist the users when integrating data sources, assisting them to see relationships within different formats.

Data visualization techniques are classified in respect of three aspects. Firstly, their focus i.e. symbolic versus geometric; secondly, their stimulus i.e. 2D versus 3D; and lastly, their display i.e. static or dynamic. The data can be presented in various visual formats including box plots, scatter plots, data distribution charts, curves, volume visualization and many others.

Advanced visualization techniques may greatly expand the range of models that can be understood by domain experts. In a data mining system, the aim of data visualization is to obtain an initial understanding of the data and the quality thereof. The actual accurate assessment of the data and the discovery of new knowledge are the tasks of the data mining tools. Therefore, the visual display should preferably be highly understandable, possibly at the cost of accuracy.

The use of one or more of the above mentioned data visualization techniques thus helps the user to obtain an initial model of the data, in order to detect possible outliers and to obtain an intuitive assessment of the quality of the data used for data mining.

2.2 Information Visualization: It is crucial to be aware of what users require for exploring data sets, small and large. The driving force behind visualizing data mining models can be broken down into two key areas, namely understanding and trust. Visualization thus aids us to determine whether the data mining process is of high economic utility i.e. it is adding value especially when considering large-scale real-world data mining projects.

The art of information visualization can be seen as the combination of three well-defined and understood disciplines, namely cognitive science, graphics art and information graphics. A number of important factors have to be kept in mind when visualizing both the execution of the data mining algorithm, e.g., the construction of a decision tree and displaying the results thereof. The visualization approach should provide an easy understanding of the domain knowledge, explore visual parameters and produce useful outputs.

The format of knowledge extracted during the mining process depends on the type of data mining task and its complexity. Examples include classification rules, association rules, temporal sequences, casual graphs etc. Visualization of these data mining results involves the presentation of the results or knowledge obtained from data mining in visual forms such as decision tree, association rules, clusters and generalized rules.

3.0 Visual Data Mining and Virtual Reality

Three-dimensional visualization has the potential to show far more information than two-dimensional visualization, while retaining its simplicity. This visualization technique quickly reveals the quantity and relative strength of relationships between elements, helping to focus attention on important entities and rules. It therefore aids both the data pre-processing and data mining processes.

Many techniques are available to visualize data in three dimensions. For example, it is very common to represent data by glyphs. Three-dimensional visualization can be made more efficient by the use of virtual reality. In traditional visualization, the human subjects looks at the data from outside, while in virtual reality the user is the part of the data world. Virtual reality is particularly well adapted to representing the scale and the topology of various sets of data.

Virtual reality can be considered as a major breakthrough in data mining. By analogy, they can be considered as the equivalent of collaborative agents in visualization. The data mining process can be carried out automatically to certain extent by distributed and collaborative agents and also collaborate on the visualization and the visual data mining aspects.

4.0 Upcoming Developments

Visual data mining is not limited to data visualization. In some specific cases, the visual appearance of the data is related to their effective functionality and visual data becomes synonym of function mining; a highly attractive features for many practical applications.

For instance, the Content-based Analysis of Protein Structure for Retrieval and Indexing (CAPRI) data mining system addresses this issue. CAPRI is able to utilize the 3D structure of a protein, in order to find the *k* most similar structures.

The main benefit of 3D structural indexing is that the protein functionality is related to its 3D shape. 3D shape indexing is a natural way to index the functionality with all the foreseen applications in bioinformatics, genomic, as well as for the pharmaceutical industry.

5.0 Conclusion

The ability to visualize the results of a data mining efforts aids the user to understand and trust the knowledge embedded in it. Data and information visualization provide the user with the ability to get an intuitive “feel” for the data and the results, e.g., in the form of rules, that is being created. This ability can be fruitfully used in many business areas, for example for fraud detection, diagnosis in medical domains and credit screening, amongst others. Finally, the direct mining of visual information looks very promising in proteomics for the design of new drugs with fewer side effects.

References

1. Blanchard J., Guillet F. and Briand H. (2006). Interactive Visual Exploration of Association Rules with Rule Focusing Methodology, *Knowledge and Information Systems*, 13 (1), 43-75
2. Garacia-Osorio C. and Fyfe C. (2008). Visualizing Multi Dimensional Data, *Successes and New Directions in Data Mining*, IGI Global, 236-276.
3. Han, J., & Kamber, M. (2006). Data Mining Concepts and Techniques, 2nd Edition, San Francisco: Morgan Kaufmann.
4. Jakun-Kelly T. J., Kwan-Liu M. and Gertz M. (2007). A Model and Framework for Visualization Exploration, *Transactions on Visualization and Computer Graphics*, 12(2), 357-369.
5. Keim D.A. (2002). Information Visualization and Visual Data Mining, *IEEE Transactions on Visualization and Computer Sciences*, 8(1), 1-8.
6. Klema J. (2007). Visualization of Multivariate Data with Tail Trees, *Information Visualization*, 6(2), 109-122.
7. Kou G., Peng Y., Shi Y. and Chen Z. (2007). Epsilon Support Vector and Large-Scale Data Mining Problems, *ICCS-2007* (Beijing, China), SLNCS 4489, 874-881.
8. Valdes J. J., Romero E. and Gonzalez R. (2007). Data and Knowledge Visualization with Virtual Reality Spaces, *IJCNN2007* (Orlando, USA), 160-165.