IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB8922

Chapter III

Cooperative Learning and Virtual Reality-Based Visualization for Data Mining

Herna Viktor
University of Ottawa, Canada

Eric Paquet
National Research Council, Canada

Gys le Roux
University of Pretoria, South Africa

ABSTRACT

Data mining concerns the discovery and extraction of knowledge chunks from large data repositories. In a cooperative datamining environment, more than one data mining tool collaborates during the knowledge discovery process. This chapter describes a data mining approach used to visualize the cooperative data mining process. According to this approach, visual data mining consists of both data and knowledge visualization. First, the data are visualized during both data preprocessing and data mining. In this way, the quality of the data is assessed and improved throughout the knowledge discovery process. Second, the knowledge, as discovered by the individual learners, is assessed and modified through the interactive visualization of the cooperative data mining process and its results. The knowledge obtained from the human domain expert also forms part of the process. Finally, the use of virtual reality-based visualization is proposed as a new method to model both the data and its descriptors.

INTRODUCTION

The current explosion of data and information, mainly caused by the extensive use of the Internet and its related technologies, e-commerce and e-business, has increased the urgent need for the development of techniques for intelligent data analysis. Data mining, which concerns the discovery and extraction of knowledge chunks from large data repositories, is aimed at addressing this need.

However, there are a number of factors that militate against the widespread adoption and use of this existing new technology in business. First, individual data mining tools frequently fail to discover large portions of the knowledge embedded in large data repositories. This is mainly due to the choice of statistical measures used by the individual tools. A number of data mining researchers and practitioners are, therefore, currently investigating systems that combine two or more diverse data mining tools. In particular, the combination of techniques that share their individual knowledge with one another is being investigated, leading to the fusion of information representing different viewpoints.

Second, the results of many data mining techniques are often difficult to understand. For example, a data mining effort concerning the evaluation of a census data repository produced 270 pages of rules (Pretorius, 2001). 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. That is, data and information visualization provides users with the ability to obtain new insights into the knowledge, as discovered from large repositories. Human beings look for novel features, patterns, trends, outliers and relationships in data (Han & Kamber, 2001). Through visualizing the data and the concept descriptions obtained (e.g., in the form of rules), 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 (Thearling, Becker, DeCoste, Mawby, Pilote & Sommerfield,2002). The user can thus get a "rough feeling" of the quality of the data, in terms of its correctness, adequacy, completeness, relevance, etc. The use of data and rule visualization thus greatly expands the range of models that can be understood by the user, thereby easing the so-called "accuracy versus understandability" tradeoff (Thearling et al., 1998).

Visual data mining is currently an active area of research. Examples of related commercial data mining packages include the *DBMiner* data mining system, *See5* which forms part of the RuleQuest suite of data mining tools, *Clementine* developed by Integral Solutions Ltd (ISL), *Enterprise Miner* developed by SAS Institute, *Intelligent Miner* as produced by IBM, and various other tools (Han & Kamber, 2001). Neural network tools such as *NeuroSolutions* and *SNNS* and Bayesian network tools such as *Hugin*, *TETRAD*, and *Bayesware Discoverer*, also incorporate extensive visualization facilities. Examples of related research projects and visualization approaches include *MLC++*, *WEKA*, *AlgorithmMatrix*, *C4.5/See5* and *CN2*, amongst others (Clark & Niblett, 1989; Fayyad, Grinstein, & Wierse, 2001; Han & Kamber, 2001; Mitchell, 1997; Quinlan, 1994) Interested readers are referred to Fayyad, Grinstein, & Wierse (2001), which provides a detailed discussion of the current state of the art.

Related Content

A Building Block Approach to Genetic Programming for Rule Discovery
Andries P. Engelbrecht, L. Schoemanand Sonja Rouwhorst (2002). *Data Mining: A Heuristic Approach (pp. 174-190).*
www.irma-international.org/chapter/building-block-approach-genetic-programming/7589

A Novel Hybrid Algorithm Based on K-Means and Evolutionary Computations for Real Time Clustering
Taha Mansouri, Ahad Zare Ravasanand Mohammad Reza Gholamian (2014). *International Journal of Data Warehousing and Mining (pp. 1-14).*
www.irma-international.org/article/a-novel-hybrid-algorithm-based-on-k-means-and-evolutionary-computations-for-real-time-clustering/116890

An UML Profile and SOLAP Datacubes Multidimensional Schemas Transformation Process for Datacubes Risk-Aware Design
Elodie Edoh-Alove, Sandro Bimonteand François Pinet (2015). *International Journal of Data Warehousing and Mining (pp. 64-83).*
www.irma-international.org/article/an-uml-profile-and-solap-datacubes-multidimensional-schemas-transformation-process-for-datacubes-risk-aware-design/130667

Ontologies in the Health Field
Michel Simonet, Radja Messaiand Gayo Diallo (2009). *Data Mining and Medical Knowledge Management: Cases and Applications (pp. 37-56).*
www.irma-international.org/chapter/ontologies-health-field/7526

A Taxonomy for Distance-Based Spatial Join Queries
Lingxiao Liand David Taniar (2017). *International Journal of Data Warehousing and Mining (pp. 1-24).*
www.irma-international.org/article/a-taxonomy-for-distance-based-spatial-join-queries/185656