Research on Visual Data Mining Technology

Kecheng Qu, Lina Wang
Shandong Xiehe University, Jinan, Shandong, P.R. China, 250109
2008xiehe@dlvte.edu.cn

Abstract. As the rapid development of computer and network technology has produced massive amounts of data, it is becoming more and more difficult for various industries to process data, and the society's demand for fast and effective data processing is also increasing. In this context, the use of traditional data processing methods is obviously insufficient. For this reason, using the advantages of visualization technology and combining it with data mining can realize the visualization of the entire process of data mining. This will not only help users understand the data mining process more clearly, increase the utilization of data mining results to ensure accurate and effective decision-making, but also improve the effectiveness and accuracy of data recognition and increase the credibility of the data mined. To this end, this article will analyze the relevant meaning and characteristics of visual data mining technology, and conduct analysis and research from several aspects such as technical design, implementation methods, application development status and future development.

1. Overview of Visual Data Mining

1.1. The Meaning of Visualization and Data Mining
Visualization technology refers to the use of computer graphics and image processing technology to realize the conversion of data information to graphic images and display them on the screen to achieve image interaction. With the development of computer technology, the application scope of visualization technology has expanded. It has become one of the main technologies used to study human-computer interaction, data expression, and data processing. The realization process of visualization technology includes data preprocessing, mapping, drawing and display. First of all, in the process of data preprocessing, for different visualization methods and contents, data compression and decompression can be carried out through reasonable settings of data formats and standards to achieve transformation processing of original data. Secondly, in the mapping stage, different mapping techniques can be used to achieve the conversion of numerical data to geometric data for different data. In the drawing stage, the conversion of geometric data to graphic images is realized. Finally, the graphic image data formed by drawing is displayed on the screen according to the user's requirements. In this process, data modeling is the core of the visualization technology.

Data mining technology refers to the mining of potentially valuable data information from a large amount of incomplete, fuzzy, and random application data. The biggest difference between data mining technology and traditional data analysis is that data mining does not have clear assumptions, while traditional data analysis first gives a hypothesis and then performs data verification. Data mining technology automatically extracts and generates a certain pattern from massive data piles through a large number of searches, and the acquired data has uncertain characteristics. The data mining object can be either an existing data file or an operational database file. The data mined through integration, cleaning,
classification, conversion and loading is guaranteed to be true and effective.

1.2. The Presentation of Visual Data Mining
The concept of visual data mining is to integrate the two technologies based on the continuous development of data mining technology and visualization technology. Visual data mining technology combines computer graphics and image processing technology to express the source data, intermediate data, final mining result model and the entire data mining process in an intuitive way for users to use, so that users can participate in the whole process of data mining expression, it is convenient to understand the whole process. The application of visual data mining technology has the following advantages. First, before mining data, use easy-to-understand graphics and images to intuitively express complex data information. This can deepen users' understanding of complex data information. Second, human-computer interaction can be used in data mining to allow users to participate in the whole process. Through the observation and management of data mining, on the one hand, the progress and depth of data mining can be accelerated, and on the other hand, the quality of data mining can be improved. Finally, the results of data mining are displayed through specific visual graphics and images. This allows users to clearly and intuitively understand data information and related knowledge and make evaluation feedback.

2. Current Status of Research and Application of Visual Data Mining Technology

2.1. Optimizing Visualization Techniques for Data Mining
Optimizing the visualization technology to realize the visualization of the whole process of data mining is the focus of current research. The improvement of visualization technology can not only express data information better and more effectively, but also realize human-computer interaction flexibly and provide reliable reference for user decision-making.

2.2. Realize the Combination of Visualization Technology and Data Mining Algorithm
The visualization research on data mining algorithms mainly uses a variety of visualization data to realize the visualization of the execution section of data mining algorithms. In this way, users can use visualization equipment and instant methods to guide and control the formulation and execution of data mining algorithms, so as to achieve the optimization of data mining visualization. In this way, the most effective combination of the two types of data can be visually described and expressed for different data mining stages. This type of visual data mining technology product should be able to support the establishment of multiple algorithms and models.

2.3. Application of Visualization in the Description of the Result Model Generated by Various Data Mining Algorithms
The research in this area is mainly to visually describe the result model generated by data mining algorithms and present it to users in visual forms such as decision trees and clustering. This allows users to make better use of the data mining result model, and helps users to visually explain and evaluate the results.

3. Design and Implementation of Visual Data Mining Technology

3.1. Design Standards
The introduction of visualization technology into the data mining process is to present the entire data mining process and results to users in a visual form. This allows users to understand the process of data mining more clearly as a whole, and to understand the interesting aspects of data mining in detail. In addition, this is also to enhance the user's sense of participation and guide the design and improvement of the human-computer interaction interface. The following factors should be fully considered in the design of visual data mining system.
First, openness. The system should be able to provide rich data access interfaces, visualization method interfaces and mapping interfaces for different data mining environments with open design standards. Second, versatility. Versatility means that it can adapt to different data mining environments, and can provide different users with a human-computer friendly and easy-to-operate interactive mode. Third, flexibility. Through a flexible management and control mechanism, control functions such as data source connection settings, data acquisition and filtering, interactive menu design and custom parameter settings are realized. At the same time, it can also perform visual data query and presentation. Fourth, scalability. Expanding the original visualization function can not only express the data mining results, but also describe the data mining process and data algorithms. Fifth, the interface is friendly. Providing a unified man-machine interface style, operation mode, etc. is also in line with user habits.

3.2. Classification of Visual Data Mining Technologies

3.2.1. Data Visualization
Data visualization refers to combining multi-dimensional data in a database from different abstract levels or combining data with different attributes and dimensions. Afterwards, a variety of visual expressions are used to present to the user, so that the user can find the relationship between the data and identify the data that needs further observation and valuable information.

3.2.2. Visualization of Data Mining Process
The visualization of the data mining process refers to the use of visualization to describe and express the entire data mining process, so that users can clearly understand how the data they need is extracted from the database and how to process the data. Even when the mining algorithm does not get the final result, it can make users understand some of the intermediate result information, so that users can participate in the data mining process, adjust and control, and find and solve problems in time. However, it is very difficult to realize the visualization of the data mining process. Therefore, we also need to strengthen research to achieve full interaction with users and ensure that the intermediate results of data mining are displayed in real time.

3.2.3. Visualization of Data Mining Result Model
The visualization of the data mining result model refers to the visualization of the results obtained by data mining in a form of visualization such as decision trees and association rules. The visualization of results can make users understand the results of data mining faster and more conveniently. In the meantime, the information obtained can be used as a reference for analysis and research to improve the accuracy and effectiveness of decision-making.

3.3. Data Visualization Technology

3.3.1. Three Elements of Data Visualization
The three elements of data visualization refer to the type of data to be visualized, the visualization method, and the technology of data interaction and transformation. Among them, the data types to be visualized are mainly standard multi-dimensional data such as one-dimensional data, two-dimensional data, and multi-dimensional display technology. Otherwise, it also includes complex data types such as hierarchical data, algorithms, and software, which can better help users understand and operate.

3.3.2. Data Visualization Technology Content
Different data types use different data visualization technologies, including chart-based visualization technology, geometric projection-based visualization technology, icon-based visualization technology, and pixel-based visualization technology. Except to these traditional data, visualization technologies based on human-computer interaction and deformation visualization technologies have emerged, which can better realize data exploration.
3.4. Visualization Technology of Data Mining Process

3.4.1. Visual Design in the Preprocessing Stage
In the data preprocessing stage, the selected data set needs to be extracted, integrated, cleaned, and converted. In this process, the use of visualization methods can give users a general understanding of the data to be processed. A large number of complex data operations are required in data preprocessing. In consequence, the interface menu operation should be combined with the traditional visual chart form in the visual design.

3.4.2. Visual Design of Data Mining Algorithm
From the perspective of algorithm visualization, the specific data mining algorithm is visually designed. Based on computer graphics, the execution process of the algorithm program and the data evolution process are vividly displayed with dynamic graphics. The visualization of specific data mining algorithms includes three functional components: algorithm initialization and related parameter setting components, intermediate result display components, and algorithm and interface mapping layer.

![Functional Components of Algorithm Visualization Design](image)

3.5. Visualization Technology of Data Mining Result Model

3.5.1. Description of Data Mining Result Model
Before the visual description of the data mining result model, a standard language can be used to express various mining data record models. For example, the latest version of PMML language uses its data layering and self-description ideas to standardize and manage data mining models.

3.5.2. Visual Design of Data Mining Result Model
The visualization of the data mining result model is to help users better understand the data mining result model and provide effective evaluation processing. There are various data mining result models, so the visual description of the data mining model is more complicated, and different data mining models need different visualization techniques. As a result, designing a model visualization interface can enable users to analyze data mining models more intuitively. The interface should include the following functional components.
4. Problems and Future Development Trends of Visual Data Mining Technology

4.1. Problems of Visual Data Mining Technology
The application that combines visualization technology and data mining technology has been greatly developed. Simultaneously, it has also played an important role in various fields. The problem of visual data mining integration has attracted the attention and research of researchers. The application of visualization to data mining is mainly to use visualization as a tool for data expression. Through visualization technology and combining the data from data mining to form a visible view, the analysis and display of some complex structure data can be realized. However, the analysis method used in this process has nothing to do with visualization technology and is the content of data mining. As a result, there is still a relatively decentralized relationship between visualization and data mining technology. We should also strengthen the connection between the two to enhance the user’s sense of participation, help end users better understand the whole process of data mining analysis and display, and realize both Functional interactivity is currently the focus of attention. Meanwhile, we should also strengthen the application of new visualization technology, further study the visualization of data preprocessing process and visualization of data mining algorithms, and improve it in continuous practice. At present, the main problems of visual data mining technology include four aspects. In terms of model building, because of the different calculation methods and the extremely fast computer speed, only the calculation results are displayed in the digital display, and the model building process is not shown. Neither the country nor the industry has formulated a universally recognized standard for measurement indicators. In addition, the problem of human-computer interaction. After the user connects the visualization device, even if they find a problem, they cannot interact and change it, which makes the visualization technology lose its original value. In the process of data mining technology and visualization technology, because the integration technology does not support the fusion of the two data under the same CPU, the connection between the two is not close enough.

4.2. The Future Development of Visual Data Mining Technology
In the future development of visual data mining, the visualization of the mining model construction process and the visualization of the data mining process should be fully realized. At the same time, we should also improve the degree and quality of visualization and actively integrate it with human-computer interaction technology. Specifically, it includes the following aspects.

First, we should strengthen the visualization technology research of unstructured and semi-structured data types, and continuously enrich the visualization of data. Second, we should actively introduce the management mechanism of the visual data mining process, continuously improve the management of the entire process of visual data mining and save operation records, so that the level of visual management of the entire data mining process is improved. Third, we need to apply the
visualization technology of the complex data mining result model, and use the visualization technology to realize the construction of the complex data structure model and the expression of data information. Fourth, we should also increase the visualization research on data mining algorithms to improve the level of interaction of data mining algorithms.

5. Conclusion
In summary, this article briefly describes the research and application status and future development trends of visual data mining technology. The author points out that integrating visualization technology and data mining technology to achieve an effective combination of all stages of the data mining process can help users to more intuitively and clearly understand the entire process of data mining analysis and presentation. In the future, we also need to increase the design update of the visual data mining technology model. We should optimize each functional component and module, and strengthen the connection between the two types of data to better promote the development and application of the visual mining technology model.

References
[1] Liu Lei. Research on Visual Data Mining Methods and Technology [J]. Communication World, 2017 (6): 296.
[2] Zeng Chunxian. Research on Visual Data Mining Technology [J]. Journal of Chongqing
[3] Gao Jian. Research on Visual Data Mining Technology [J]. Communication Design and Application, 2018 (1): 116-117.