

Researches on the Prototype Implementation of Visual Data Mining Techniques

Guiliang Feng, Zhonghua Li and LianChun Zhang*

*HeBei North University, China
6838710@qq.com*

Abstract

This paper describes the core functional components and a visualization module implementation process of visual data mining technology prototype system. The first part introduces the implementation technology of visual data mining technology prototype system and then describes the overall design of the architecture and features of visual data mining technology prototype system; Then paper describes the common components in the visualization of data mining technology prototype system. Research achievements of this article provides a useful reference to optimize visual data mining techniques.

Keywords: *visual data mining technology; implementation techniques; Architecture*

1. Introduction

Visual data mining techniques generates with the development of data mining technology and information visualization technique, and it effectively applied human perception and knowledge in the field of data mining process. It is based on human perception patterns, exceptions, aptitude and relationships by visualization to enhance data mining processing. Data mining is a process of extracting a potentially valuable knowledge from large amounts of historical data.

Visual data mining technology combines the processing power of computers, people's initiative and the intuitive nature of data, expanding multi-dimensional fort the information flow. In traditional data mining technology, the user cannot observe the data mining process, and the process does not have the intuitive. Visual data mining technology is closely linked to computers and users, to provide users with an intuitive information to facilitate user to interact flow data, greatly improving the efficiency of data mining and it obtained more valuable data results.

Visualization is a process that data, information and knowledge representation is converted to visible, which provides an interface between human and computer information processing systems. The use of effective visualization techniques can quickly and efficiently deal with large amounts of data to find the hidden features, relationships, patterns and trends that can guide a new predictable and more efficient decision-making. Some data mining techniques and algorithms make decision-makers difficult to understand and use. Visualization can make it easier to understand the mining results, and allows the comparison of test results, also used to guide data mining algorithms, allowing users to participate in the process of decision analysis.

2. The Implementations of the Main Function Module

2.1 Implementations of the Common Components

2.1.1 Implementations Data Access Interface

In the text file access interface via Java file read operation, the CSV file is read into memory and converted to internal data format. Because the text file format accessible through a text file interface needs to meet certain format specifications to divide the data records and behavior records with a specific delimiter. In this case, the text file interface is mainly supported by the text file access of generic CSV format. Here it requires a text file in CSV format storages column names in the first line, commas are made of the data record delimiter and do not cross a line of data. In addition to the local file data, interface can also load a text file CSV file which is specified by HTTP and FTP path. JDBC interface obtains data from a local/network data source to provide a data visualization module, or get the data from the data preprocessing stage of pre-processing component data mining system, providing visualization module of data mining process.

Because JDBC is independent of the specific database system, it is possible to provide a uniform access to a variety of relational databases, and without changing the program to access the database when the database is changed, here, the data access component packages several common steps for JDBC database access, implementing component to reuse. JDBC package type is shown in Figure 1. Through the above type of data access, database connectivity step may consist of the following steps:

- 1) get connected: `Connection conn=DriverManager.getConnection(ConnURL);`
- 2) execute SQL: `Statement. executeQuery(); Statement. executeUpdate ();`
- 3) results obtained: `ResultSet. next (); ResultSet. getXXX();`
- 4) Exception Handling: `try { ... } catch (SQLException e) { ... }`
- 5) Close the connection: `conn. close ();`

2.1.2 Implementations of Data Memory Interface

Because the data access components get data from local / network files and text files in a relational database and source data has a different data formats, therefore, it is necessary to define a unified data memory interface, the data object from different data sources are converted into a single data object, then subjected to treatment, or other visual access operation. The data memory interface shown in Figure 2.

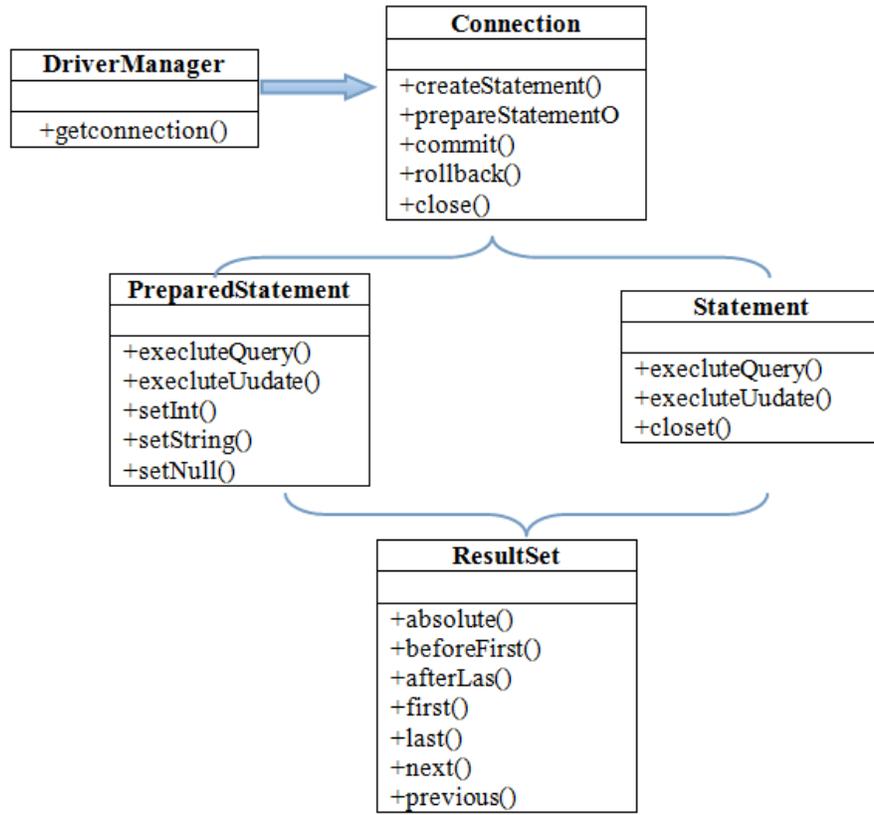


Figure 1. Data Access Package Class Diagram

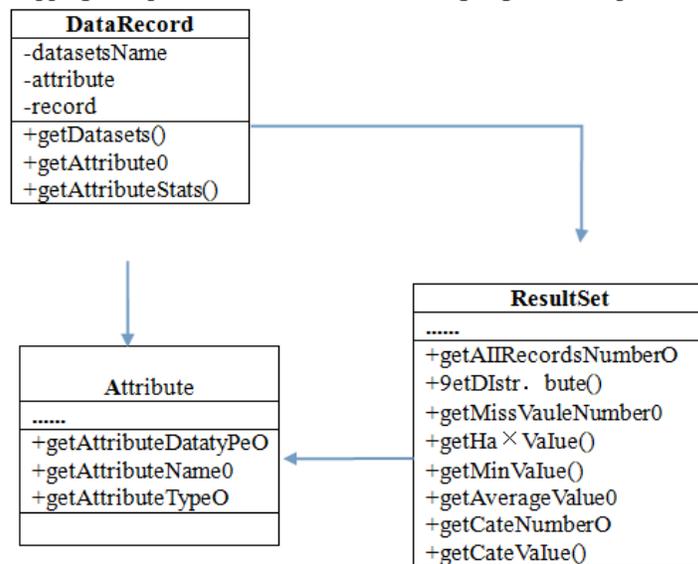


Figure 2. Data Storage Interface

2.2 Implementations of the Mapping Component

Mapping component Associated data mining algorithm implementation and interface operation, providing a uniform interface for visualization algorithm. This article package data mining process and the associated data mining algorithms through data mining DMX extension language mainly using CREATE, INSERT, SELECT these three DMX statement. Table 1 shows the syntax of the DMX statements.

Table 1. Part DMX Sentence Structure

DMX statement	Syntactic structure
CREATE	CREATE[SESSION[MINING MODEL<data mining model name> ([(<a list of data mining algorithms used, separated by commas>)]) USING<data mining algorithms>[(<parameter List>)]
INSERT	INSERT INTO MINING MODEL<data mining model name> (<column names, separated by commas>)<Getting SQL data source>
SELECT	SELECT<column names, separated by commas> FROM<data mining model name>.CONTENT [WHERE<query conditions>] [ORDER BY<sort condition>[DESCIASC]]

CREATE statement is used to create patterns of data mining process, by creating a data mining model table in the database to store the data mining models associated with the information, such as the use of the algorithm, the algorithm parameters.

INSERT statements is used for training data. Data mining model based on data mining model name to find the data defined in the database, using the data mining algorithms and parameters, and then as a data mining algorithms with the specified input source data, the final training algorithm outputs the result will be stored in the data mining results model.

SELECT statement is used to query or data mining models to predict obtain intermediate results and information in the process of data mining algorithms executed.

2.3 Implementations of MML Analytical Components

PMML analytical component provides interfaces to resolve a variety of data mining model results PMML document, PMML document with tissue forms of XML documents in Java, it can be parsed XML documents through SAX AP. It uses an event-based model, the basic principle is the sequential scan for XML documents when scanning a document the beginning and end, beginning and end elements such places, notice in event handler, take appropriate action by the event handler. Specific comprising:

startDocument/endDocument: Trigger document processing events at the beginning and end of the document.

startElement/endElement: Each element in the document being parsed before and after trigger element event.

Characters: Generating DTD or Schema events. In the processing of the document DTD or Schema.

In addition, SAX notify the host application to appear parsing error by error events. The advantage of using Java SAX to parse XML document is when SAX parsing XML documents, without waiting for all the contents have been read before starting treatment. It began when reading the document inspection data, the other relatively Java parsing XML documents, SAX parsing much faster. Therefore, this paper PMML analytical components uses Java SAX parser to parse different data mining results described in PMML models.

3. Implementations of Data Visualization Function Modules

Data visualization module is a basic module in the visual data mining techniques prototype system. Its function is: Before data mining, allowing users to get a general understanding of the distribution and related statistical information of the source data, then take some pre-processing operations based on the results obtained after the analysis, such as data type conversion, data cleaning. In implement of this module, this paper considers the discrete and continuous visualization of these two types data, using visualization techniques based on the chart, and achieved through the Java Swing component technology and open source JFreeChart chart library.

3.1 JFreeChart Chart Library Source

JFreeChart chart library source code is mainly composed of two major packages: org.jfree.chart and org.jfree.data. The former is mainly concerned with the graphic itself, the latter data is relating to the graphical display. This paper uses the core classes of these two packages, as shown in Table 2:

Table 2. JFreeChart Core Classes

org.jfree.chart.JFreeChart	JFreeChart engine provides a factory class ChartFactory, for creating different types of chart objects
org.jfree.data.category.XXXDataSet	Dataset object provides the data used for the chart display. Different types of charts corresponding to different types of data sets object classes, such as PieDataset, categorydataset, etc.
org.jfree.chart.plot.XXXPlot	Chart area object, chart styles, such as the vertical axis and the horizontal axis shows the label display
org.jfree.chart.axis.XXXAxis	Two vertical and horizontal axis of the graph processing
org.jfree.chart.renderer.XXXRender	Shows the generated chart object
XXXToolTipGenerator:	Generate image help tips, different types of graphs corresponding to different types of tool tip category

3.2 Event Listener Mechanism

Event listener mechanism is used switch to a different visualization chart shows. Depending on the selected property or graphically, Where in AttributesListPanel, add an event listener for JList component object to display the corresponding property visualization charts, based on the properties of the currently selected, J is shown in Figure 3.

```
// lenseAttrib is a List instance
lenseAttrib.addItemListener( new ItemListener(){
    public void itemStateChanged(ItemEvent ie){
        if(ie.getStateChange() == ItemEvent.SELECTED){
            classIndex=lenseAttrib.getSelectedIndex()-1;
            if(classIndex !=null){
                Attribute attribute=getAttribute(classIndex);
                // Select the new property from the property list,Re-initialize
                setAttributeData(Attribute attribute);
            }
        }
    }
});
```

Figure 3. List Instance Object Event Listener Pseudo Code

4. Implementations of Visualization Functions Module in the Data Mining Process

Data preprocessing stage includes these five steps, that is, data extraction, data integration, data cleansing, data conversion and data protocol. In the data pre-process, When the source data is imported through data extraction, data mining will be stored in the system database or data warehouse data. Any operation in a copy of the pre-processing stage, and is stored for the corresponding changes in the source data for the instructions from the sub. The main purpose of the introduction of the visualization process data preprocessing is to provide users with editable data tables. Visual data mining prototype technology to achieve chart display and auxiliary information to guide the user to edit the data, fill in missing values, data format conversion and other operations. This paper mainly uses Java Swing component technology to build a data pre-processing wizard proposing development of the various steps envisaged in the data preprocessing.

This article uses association rules algorithm to study visualization of data mining algorithms, the algorithm can be effective requires interaction with the user interface during the operation, which requires high efficiency of the algorithm. Therefore, in the implementation process, this paper connected an improved Apriori algorithm. By splitting the transaction log data, split each transaction record data until the smallest item, and then get a single transaction log data for all projects set combination, so when finished reading the entire database, transaction log data will all subsequently splitting completed and can get all the items set combination. The improved algorithm can extract the two parameters of minimum support and minimum confidence, allowing the user to set the parameters of these two interfaces dynamically generate the required frequent item sets and association rules thereby supporting the interaction with the user in data mining association rules visualization environment.

5. Implementations of the Visualization Functions Module of Data Mining Model Results

5.1 PMML Description of the Data Mining Model Results

This article uses PMML to describe and store the results of data mining, so that it can seamlessly exchange or share data mining results model with other data mining software products/tools, because PMML is an XML-based document organization, this article uses API

provided by JAVA dom4j to build a PMMLcerator class, which provides a method for generating a corresponding PMML document data mining results for different models. Thus the data mining results are stored as PMML model file format.

5.2 The Graphical Visualization of Data Mining Model Results

The graphical visualization of data mining results model uses a combination of the chart to show, the results of different data mining models correspond to different models observer. Each model PMML analytical observer reuse components, interfaces to achieve the corresponding analytical results for different data mining model PMML document. Association rules displayed to the user through the JTable component, and users can filter, sort rules to see the interest through the interface. Meanwhile, users can select the appropriate association rules to import a text file for subsequent analysis required. In addition the form displays the association rules, the API provided by universal graphics library JUNGI constructed a visual chart class of data mining model results, providing a common graphical and mode of operation of these graphics.

6. Conclusion

Visual data mining technology prototype developed in this chapter provides the basic functions needed for visual data mining, and data mining can be used as experimental platform and take advantage of the study. With the continuous development of visual data mining techniques, the application of visual data mining techniques are increasingly being used. Our country is only recently begun to engage in researching in this area this year, this technology is not yet perfect, there are many issues need to be resolved.

Acknowledgements

This work was supported by the Hebei North University key project "based on cloud computing platform of student work early warning system research", project number:ZD201303, corresponding author: Zhang Lianchun.

References

- [1] C. Biao, "Visualization of data mining technology thinking", *Ecommerce*, vol. 9, (2013), pp. 67-69.
- [2] Z. M. Lai Lin, "Applied research techniques of visual data mining", *Digital technology and applications*, (2010), pp. 11-22.
- [3] N. Shenglin, "Relationship management based on data mining commercial banking customers", *Institute of Fiscal Science*, (2013).
- [4] Z. Xiaomei, W. Qianping and S. Lin, "XML-based Web data mining design model", *Computer Engineering and Design*, (2007), pp. 272 -277.
- [5] D. Wei, "Visualization of spatial clustering algorithm for mining research and application", *Jilin Agricultural University*, (2012).
- [6] T. Yuan, "Data Mining Technology Academic System Design and Implementation", *University of Electronic Science and Technology*, (2012).

Authors



Feng Guiliang, lecturer of department of information science and engineering, Hebei North University, He is good at the field of software engineering and multimedia development.



ZhongHua Li, her primary research interests are Data Mining, Big Data Computing and Algorithms along with their applications in several domains.



LianChun Zhang, he is a professor at Hebei North University, author for correspondence.