

Data Mining - Applications & Trends

Introduction

Data Mining is widely used in diverse areas. There are number of commercial data mining system available today yet there are many challenges in this field. In this tutorial we will applications and trend of Data Mining.

Data Mining Applications

Here is the list of areas where data mining is widely used:

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific Applications
- Intrusion Detection

FINANCIAL DATA ANALYSIS

The financial data in banking and financial industry is generally reliable and of high quality which facilitates the systematic data analysis and data mining. Here are the few typical cases:

- Design and construction of data warehouses for multidimensional data analysis and data mining.
- Loan payment prediction and customer credit policy analysis.
- Classification and clustering of customers for targeted marketing.
- Detection of money laundering and other financial crimes.

RETAIL INDUSTRY

Data Mining has its great application in Retail Industry because it collects large amount data from on sales, customer purchasing history, goods transportation, consumption and services. It is natural that the quantity of data collected will continue to expand rapidly because of increasing ease, availability and popularity of web.

The Data Mining in Retail Industry helps in identifying customer buying patterns and trends. That leads to improved quality of customer service and good customer retention and satisfaction. Here is the list of examples of data mining in retail industry:

- Design and Construction of data warehouses based on benefits of data mining.
- Multidimensional analysis of sales, customers, products, time and region.
- Analysis of effectiveness of sales campaigns.
- Customer Retention.

- Product recommendation and cross-referencing of items.

TELECOMMUNICATION INDUSTRY

Today the Telecommunication industry is one of the most emerging industries providing various services such as fax, pager, cellular phone, Internet messenger, images, e-mail, web data transmission etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business.

Data Mining in Telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service. Here is the list examples for which data mining improve telecommunication services:

- Multidimensional Analysis of Telecommunication data.
- Fraudulent pattern analysis.
- Identification of unusual patterns.
- Multidimensional association and sequential patterns analysis.
- Mobile Telecommunication services.
- Use of visualization tools in telecommunication data analysis.

BIOLOGICAL DATA ANALYSIS

Now a days we see that there is vast growth in field of biology such as genomics, proteomics, functional Genomics and biomedical research. Biological data mining is very important part of Bioinformatics. Following are the aspects in which Data mining contribute for biological data analysis:

- Semantic integration of heterogeneous , distributed genomic and proteomic databases.
- Alignment, indexing , similarity search and comparative analysis multiple nucleotide sequences.
- Discovery of structural patterns and analysis of genetic networks and protein pathways.
- Association and path analysis.
- Visualization tools in genetic data analysis.

OTHER SCIENTIFIC APPLICATIONS

The applications discussed above tend to handle relatively small and homogeneous data sets for which the statistical techniques are appropriate. Huge amount of data have been collected from scientific domains such as geosciences, astronomy etc. There is large amount of data sets being generated because of the fast numerical simulations in various fields such as climate, and ecosystem modeling, chemical engineering, fluid dynamics etc. Following are the applications of data mining in field of Scientific Applications:

- Data Warehouses and data preprocessing.
- Graph-based mining.
- Visualization and domain specific knowledge.

INTRUSION DETECTION

Intrusion refers to any kind of action that threatens integrity, confidentiality, or availability of network resources. In this world of connectivity security has become the major issue. With increased usage of internet and availability of tools and tricks for intruding and attacking network prompted intrusion detection to become a critical component of network administration. Here is the list of areas in which data mining technology may be applied for intrusion detection:

- Development of data mining algorithm for intrusion detection.
- Association and correlation analysis, aggregation to help select and build discriminating attributes.
- Analysis of Stream data.
- Distributed data mining.
- Visualization and query tools.

Data Mining System Products

There are many data mining system products and domain specific data mining applications are available. The new data mining systems and applications are being added to the previous systems. Also the efforts are being made towards standardization of data mining languages.

Choosing Data Mining System

Which data mining system to choose will depend on following features of Data Mining System:

- **Data Types** - The data mining system may handle formatted text, record-based data and relational data. The data could also be in ASCII text, relational database data or data warehouse data. Therefore we should check what exact format, the data mining system can handle.
- **System Issues** - We must consider the compatibility of Data Mining system with different operating systems. One data mining system may run on only on one operating system or on several. There are also data mining systems that provide web-based user interfaces and allow XML data as input.
- **Data Sources** - Data Sources refers to the data formats in which data mining system will operate. Some data mining system may work only on ASCII text files while other on multiple relational sources. Data mining system should also support ODBC connections or OLE DB for ODBC connections.
- **Data Mining functions and methodologies** - There are some data mining systems that provide only one data mining function such as classification while some provides multiple data mining functions such as concept description, discovery-driven OLAP analysis, association mining, linkage analysis, statistical analysis, classification, prediction, clustering, outlier analysis, similarity search etc.
- **Coupling data mining with databases or data warehouse systems** - Data mining system need to be coupled with database or the data warehouse systems. The coupled components are integrated into a uniform information processing environment. Here are the types of coupling listed below:
 - No coupling
 - Loose Coupling
 - Semi tight Coupling
 - Tight Coupling
- **Scalability** - There are two scalability issues in Data Mining as follows:
 - **Row (Database size) Scalability** - Data mining System is considered as row scalable when the number or rows are enlarged 10 times, It takes no more than the 10 times to execute the query.

- **Column (Dimension) Scalability** - Data mining system is considered as column scalable if the mining query execution time increases linearly with number of columns.
- **Visualization Tools** - Visualization in Data mining can be categorized as follows:
 - Data Visualization
 - Mining Results Visualization
 - Mining process visualization
 - Visual data mining
- **Data Mining query language and graphical user interface** - The graphical user interface which is easy to use and is required to promote user guided, interactive data mining. Unlike relational database systems data mining systems do not share underlying data mining query language.

Trends in Data Mining

Here is the list of trends in data mining that reflects pursuit of the challenges such as construction of integrated and interactive data mining environments, design of data mining languages:

- Application Exploration
- Scalable and Interactive data mining methods
- Integration of data mining with database systems, data warehouse systems and web database systems.
- Standardization of data mining query language
- Visual Data Mining
- New methods for mining complex types of data
- Biological data mining
- Data mining and software engineering
- Web mining
- Distributed Data mining
- Real time data mining
- Multi Database data mining
- privacy protection and Information Security in data mining

Source:

http://www.tutorialspoint.com/data_mining/dm_applications_trends.htm