JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DATA MINING AND DATABASES

Prof. Sandeep Kaur

Assistant Professor in Computer Science

Miri Piri Khalsa College, Bhadaur (Barnala)

ABSTRACT:

Mining databases and data processing is crucial in today's information technology. This paper is concerned with the mining and databases, their relationships, and methods employed in it along with its challenges and applications. With the rapid increase in data, it is important to know how the data gets stored, managed, and preprocessed in the databases for data mining purposes. This paper further examines the forthcoming trends and the effects of these technologies on industries and research.

1. INTRODUCTION

Data mining also known as knowledge discovery in databases (KDD) refers to the process of discovering patterns and trends, as well as meaningful pieces of information and insights from large datasets. Depending on the type of document, it can be divided further into sections. For effective data mining, the suitable tools and techniques need to be applied in order to make use of the insights at hand.

Due to the extensive use of the internet, social media, IoT and other technologies, there has been a rapid increase in the rate of data generation and because of which, the importance of data mining has also increased significantly. In this paper, we discuss:

- Interrelation of data mining and databases
- Methods of data mining
- Problems faced and their corresponding applications
- Future developments of the area

2. DATABASES: THE DEVELOPMENT OF DATA MINING

A database is a collection of data in a format that enables easy access, management, and updating. With time, the development of databases has undergone a metamorphic change moving from simple flat files to the more advanced relational databases (RDBMS) to the more recent non-relational or NoSQL databases.

2.1 TYPES OF DATABASES

- Relational Databases (RDBMS): These databases use SQL as a standard programming language for managing data. Examples are MySQL and PostgreSQL.
- NoSQL Databases: These databases are used for semi-structured and unstructured data. Examples are MongoDB and Cassandra.
- Data Warehouses: These databases are specialized in analytical queries along with data aggregation. Examples are Amazon Redshift and Snowflake.
- Distributed Databases: These databases are capable of distributing data across a particular collection of servers.

2.2 IMPLICATIONS IN DATA MINING

Miner jobs are concerned with mining tasks of raw data that is supplied from these systems. Good quality database systems make sure that data is available in the required formats, which is the first step in improving successful data mining efforts.

3. DATA MINING: TECHNIQUES AND PROCESSES

Data mining is the practice of analyzing data from different perspectives and summarizing it into useful information. In these processes, the knowledge obtained through the data is put together with the use of statistics, machine learning, and computer science databases.

3.1 KEY TECHNIQUES

- Classification: Assigning data into specific categories provided. Algorithms: Decision Trees, Naive Bayes.
- Clustering: Grouping of similar data points. Algorithms: K-Means and DBSCAN.
- Association Rule Mining: Learning the relationship between different variables. An example is the apriori algorithm for market basket analysis.

- Regression: A form of analysis that predicts the new continuously changing value depending on how it was previously defined.
- Anomaly Detection: Remembering unusual data points unlike the rest.

3.2 KNOWLEDGE DISCOVERY IN DATABASES (KDD) PROCESS

- Data Cleaning: Getting rid of inconsistencies and noise.
- Data Integration: Joining data from more than one source together.
- Data Transformation: Modifying data to prepare it for analysis.
- Data Mining: Using algorithms to pull out useful pattern information.
- Pattern Evaluation: Ensure the mined patterns validate.
- Knowledge Presentation: Results visualization in a way that makes interpretation easy.

4. GAPS AND ISSUES ENCOUNTERED IN MINING AND DATABASES

- Data's Validity: Assessing incomplete or inaccurate data can render analysis useless.
- Scaling: The issues associated with data mining and management of huge volumes of data.
- Privacy of Data: The ethical use of data vs data privacy.
- Varied Data Sources: Alignment of structured, semi-structured, and unstructured data.
- Data Stream Processing: The effectiveness of managing streaming data.

5. USES OF DATA MINING AND DATABASE

5.1 BUSINESS

- Customer Relationship Management (CRM): Evaluating current CRM to increase customer satisfaction.
- Market Basket Analysis: Understanding buying habits and adjusting stock levels.

5.2 HEALTHCARE

- Monitoring patient records to forecast potential epidemics.
- Looking for relationships in data concerning genes for medical treatment.

5.3 FINANCE

- Monitoring purchase activity on credit accounts or cards to identify fraudulent activity.
- Loan or investment evaluation.

5.4 EDUCATION

- Tracking student data to establish whether they can cope with challenges presented to them.

5.5 SOCIAL MEDIA

- Public opinion market research.
- Finding topics or trends of interest.

6. MERGING DATA MINING INTO DATABASES

To render data mining more effective, the modern day databases embed data mining functions. These are the examples:

- SQL-Based Data Mining: Incorporating SQL with data mining capabilities, for example Microsoft SQL Server Analysis Services (SSAS).
- In-Database Analytics: To save on data transmission costs, data mining work is done on the databases themselves.
- Big Data Frameworks: Mining data using Hadoop or Spark tools on a large scale.

7. UPCOMING TRENDS

- Automated Data Mining: Creating AI tools for self-sufficient data pattern recognition.
- Graph Databases: For storing and mining complex network relationships such as Neo4j.
- IoT Integration: Real-time data mining through sensor data analysis.
- Quantum Mining: Mining activities sped up by quantum algorithms.
- Responsible Data Mining: Practicing ethical ways of handling data.

8. CONCLUSION

The two concepts of data mining and databases are interdependent, each contributing to and complementing the expansion of the amount of data present in the world powered by technology in the present world. Although there are issues that are unaddressed, the evolution of processes and technologies are setting the positive direction for the future of these fields. The advanced data mining methods integrated into sophisticated database systems will enable the industry and academic researchers to effectively utilize data for technology and decision making in innovative ways.

REFERENCES

1. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Elsevier.
2. Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, 13(6), 377-387.
3. Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. Proceedings of the 1993 ACM SIGMOD.
4. Zaki, M. J., & Meira Jr, W. (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press.
5. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. ACM Computing Surveys, 41(3), 1-58.
6. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. AI Magazine, 17(3), 37-54.