



INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS

ISSN 2320-7345

TRENDS OF MINING TECHNIQUES, APPLICATIONS AND TOOLS: A SURVEY

S.Vasanthi¹, S.Anusuya Devi²

¹S.Vasanthi, Assistant Professor, Department of Computer Science, PSG College of Arts and Science, Coimbatore, Tamilnadu, India, E-mail: sn_vasanthi@yahoo.co.in

²S.Anusuya Devi, Research scholar, Department of Computer Science, PSG College of Arts and Science, Coimbatore, Tamilnadu, India, E-mail: anujenifers@gmail.com

Abstract

In contemporary world all industries consists of huge amount of data. There is no use until it to be converted as useful information. In this paper we have focused a variety of Mining, techniques, approaches and different areas of the research which are helpful and marked as the important field of advanced data mining Technologies. Although this review cannot claim to be exhaustive, it does provide reasonable insights and shows the incidence of research on this subject. This paper imparts more number of applications of the data mining which will helpful in the further research. Data Mining is widely used in various areas, such as Bio-Informatics, Medial, Games, Human Rights, Banking and Finance, Business Intelligence, Retail and On-Line shopping.

Keywords: Data Mining, Spatial Mining, Image Mining, Text Mining, Visual Mining

1. Introduction

Data mining is a technique that has huge applications in a wide variety of many fields. It is used by businesses or organizations, Health care; spatial data that need to recognize certain patterns or trends. Data mining is a direct result of the increasing use of computer databases in order to store and retrieve data in repository. Highly advanced computers may be capable of predicting the best treatment for a disease. Cosmologists could use data mining to help them find out more information about the nature of the universe. Data mining may have a technological impact that could even be compared to the internet itself.

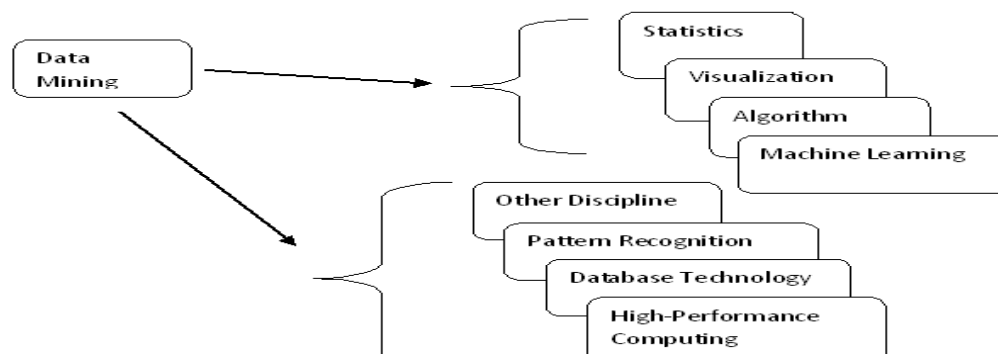


Fig 1 Data Mining Techniques

1.1 Trends of Data Mining

- ✓ Application exploration: Dealing with application-specific problems
- ✓ Scalable and interactive data mining methods
- ✓ Integration of data mining with Web search engines, database systems, data warehouse systems and cloud computing systems
- ✓ Mining social and information networks
- ✓ Mining spatiotemporal, moving objects and cyber-physical systems
- ✓ Mining multimedia, text and web data
- ✓ Mining biological and biomedical data
- ✓ Data mining with software engineering and system engineering
- ✓ Visual and audio data mining
- ✓ Distributed data mining and real-time data stream mining
- ✓ Privacy protection and information security in data mining

1.2 Data mining Tools

- Environment for Developing KDD-Applications Supported by Index-Structures (ELKI) — data mining software framework written in Java with a focus on clustering and outlier detection methods.
- SCAViS — Java-based data analysis framework
- Konstanz Information Miner (KNIME)
- OpenNN — Open source neural networks software library written in the C++ programming language.
- Orange (software) — data visualization and data mining for novice and experts, through visual programming or Python scripting. Extensions for bioinformatics and text mining.
- Rapid Miner — data mining software written in Java, fully integrating Weka, featuring 350+ operators for pre-processing, machine learning, visualization, etc.
- Scriptella ETL — ETL (Extract-Transform-Load) and script execution tool. Supports integration with J2EE and spring. Provides connectors to CSV, LDAP, XML, JDBC/ODBC and other data sources.
- Weka — data mining software written in Java featuring machine learning operators for classification, regression, and clustering.

- Jasper Soft — data mining with programmable abstraction layer.

2. Spatial Mining

Spatial data mining is the application of data mining methods to spatial data. The objective of spatial data mining is to find patterns in data with respect to geography. So far, data mining and Geographic Information Systems have existed as two separate technologies, each with its own methods, traditions, and approaches to visualization and data analysis. Particularly, most contemporary GIS have only very basic spatial analysis functionality. The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasize the importance of developing data-driven inductive approaches to geographical analysis and modelling. The main difference between data mining and spatial data mining is that in spatial data mining tasks we use not only non-spatial attributes, but also spatial attributes.

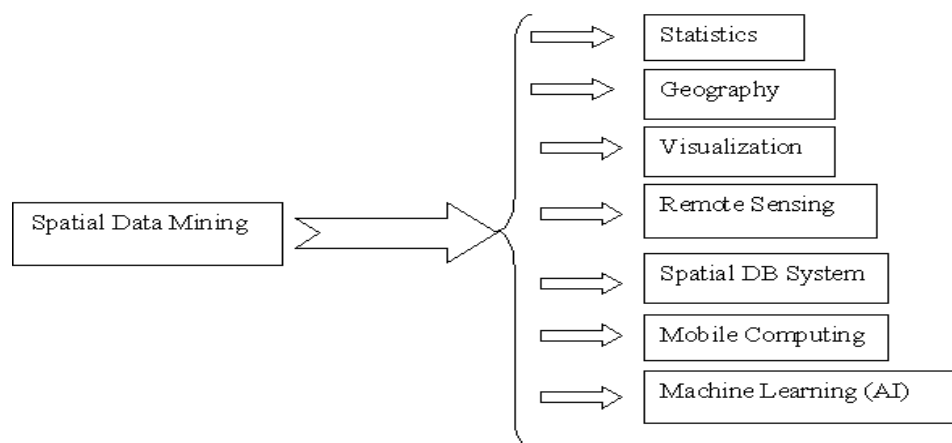


Fig 2: Spatial Mining Techniques

2.1 Spatial Mining Tools

IRedistrict:

Redistricting Software with Human-Computer Collaboration: IRedistrict is patent-pending redistricting software, with powerful optimization algorithms and intuitive user control.

EntroMap:

Detecting Spatially Varying Multivariate Relationships: EntropMap is a new approach that can detect the existence of multivariate relationships without assuming a prior relationship form.

Flow Mapping with Graph Partitioning and Regionalization:

Flow Mapping with Graph Partitioning and Regionalization is an integrated software tool to explore flow patterns in large spatial interaction data.

REDCAP:

Regionalization with Constrained Clustering and Partitioning: Regionalization is to divide a large set of spatial objects into a number of spatially contiguous regions while optimizing an objective function, which is normally a homogeneity measure of the derived regions.

VIS-STAMP:

A Visualization System for Space-Time and Multivariate Patterns: VIS-STAMP is a geo-visual analytic software package that couples computational, visual, and cartographic methods for exploring and understanding spatio-temporal and multivariate data. It can help analysts investigate complex patterns across multivariate, spatial, and temporal dimensions via clustering, sorting, and visualization.

SOMVIS Multivariate Mapping and Visualization:

SomVis is an integrated software tool that is able to: (1) perform multivariate analysis, dimensional reduction, and data reduction with the Self-Organizing Map (SOM); (2) encode the SOM result with a systematically designed color scheme; (3) visualize the multivariate patterns with a modified Parallel Coordinate Plot (PCP) display and a geographic map (GeoMap); and (4) support human interactions to explore and examine patterns.

POLO: Program for Optimal Linear Ordering

Linear ordering is to transform a set of data objects to a one dimensional order, which preserves multivariate or graph patterns as much possible. Linear ordering is widely used to discover patterns. When applied in geographical analysis, a linear ordering can be used to facilitate the detection of space-time, space-multivariate, and spatial interaction patterns.

3. Image Mining

Image mining is the process of searching and discovering valuable information and knowledge in large volumes of data. Image mining draws basic principles from concepts in databases, machine learning, statistics, pattern recognition and 'soft' computing. Using data mining techniques enables a more efficient use of data banks of earth Observation data. It is thus becoming an emerging research field in geosciences because of the increasing amount of data which lead to new promising applications. Using very high resolution satellite images now enables the observation of small objects, while the use of very high temporal resolution images enables monitoring of changes at high frequency.

3.1 Image Mining Tools

CellAnalyst is an advanced image-mining tool in an exceptionally user-friendly package. Cell Analyst is designed to automate the analysis of digital images. It identifies cells within the image, computes their statistics, and saves the data for further processing.

SmutDetect for Linux v.0.5.03-Beta: - ***SmutDetect*** is skin-tone image mining software which scans directories for images containing a specified percentage of skin-tones.

Pixcavator Scientific Image Analysis v.2.4:- Pixcavator provides new image analysis capabilities to scientists and researchers.

Bersoft Image Measurement v.4.01:- Bersoft Image Measurement (BIM) is designed as a flexible tool for acquiring, measuring and analyzing digital images.

Image Converter .EXE v.2.0.46:- Image Converter .EXE is a fast and flexible image conversion program, which allows support for converting to over 15 formats. It features over 20 effects which can be applied in any conversion process, and even in batch mode.

Acronis True Image Enterprise Server v.8.0:- Acronis True Image Enterprise Server for Windows allows to create the exact server disk image for complete server backup providing organizations with the most comprehensive, reliable, and cost effective server protection.

DC Image Button v.2.1:- DC Image Button is the only program that gives you full control over your application's interface. Give your program a creative boost by adding your own graphic buttons, icons, and images.

4 .Text Mining

Text mining is the analysis of data contained in natural language text. Text mining works by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques. Information can be extracted to derive summaries for the words contained in the documents or to compute summaries for the documents based on the words contained in them. Text mining can help an organization derive potentially valuable business insights from text-based content such as word documents, email and postings on social media streams like Face book, Twitter and LinkedIn.

Mining unstructured data with natural language processing, statistical modeling and machine learning techniques can be challenging, however, because natural language text is often inconsistent. It contains ambiguities caused by inconsistent syntax and semantics, including slang, language specific to vertical industries and age groups, double entendres and sarcasm. Text analytics software can help by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques. With an iterative approach, an organization can successfully use text analytics to gain insight into content-specific values such as sentiment, emotion, intensity and relevance.

4.1 Text mining Applications

The technology is now broadly applied for a wide variety of government, research, and business needs. Applications can be sorted into a number of categories by analysis type or by business function. Using this approach to classifying solutions, application categories include:

- E-Discovery, Records Management
- National Security/Intelligence
- Scientific discovery, especially Life Sciences
- Sentiment Analysis Tools, Listening Platforms
- Natural Language/Semantic Toolkit or Service
- Publishing
- Automated ad placement
- Search/Information Access
- Social media monitoring
- Enterprise Business Intelligence/Data Mining, Competitive Intelligence

4.2 Text Mining Techniques

- ✓ Information Retrieval
- ✓ Information Extraction
- ✓ Web Mining
- ✓ Clustering

4.3 Text Mining Tools

AeroText: – a suite of text mining applications for content analysis. Content used can be in multiple languages.

Mathematica: - provides built in tools for text alignment, pattern matching, clustering and semantic analysis.

Medallia: - offers one system of record for survey, social, text, written and online feedback.

Megaputer Intelligence: - offers text and data mining and analytics - developers of PolyAnalyst software.

SAS: – SAS Text Miner and Teragram; commercial text analytics, natural language processing, and taxonomy software used for Information Management.

Rapid Miner with its Text Processing Extension – data and text mining software.

5. Multimedia Mining

Multimedia Data Mining is the sub branch of Data Mining. Multimedia Data mining further classified according to the multimedia data type that are image, audio, video. Mostly, Multimedia data mining covers the mining of image data, mining of audio data and mining of video data.

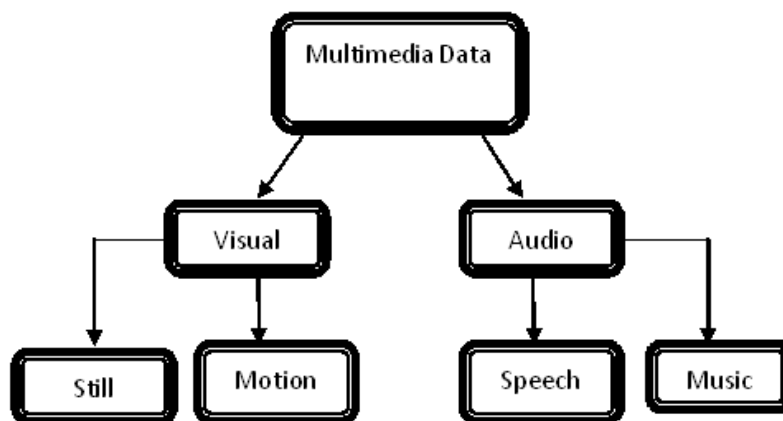


Fig 3: Multimedia Data

5.1 Multimedia Mining Applications

- ✓ Musical copyright resolution
- ✓ Chorus and pattern identification in songs
- ✓ Traffic monitoring systems

- ✓ Speaker emotion recognition in audio
- ✓ Automatics summarization of TV programs
- ✓ Sound effect Retrieval
- ✓ Musical information Retrieval System

5.2 Multimedia Mining Tools

➤ *MultiMediaMiner*

The Image Excavator and the pre-processor have been enhanced to collect and pre-process more information necessary for the MultiMediaMiner.

➤ *Rapid Miner*

Rapid Miner is easily the most powerful and intuitive graphical user interface for the design of analysis processes. An easy-to-use visual environment allows you to recognize errors, apply quick fixes, and see quick and accurate results — without needing to sift through code.

6. Visual Mining

Visual Data Mining (VDM) is a new approach for exploring very large data sets, combining traditional mining methods and information visualization techniques. Visual data mining “discovers implicit and useful knowledge from large data sets using data and or knowledge visualization techniques”. The three definitions recognize that visual data mining relies heavily on human visual processing channel, and utilizes human cognition. The three definitions also emphasis, respectively, the key importance of the following three aspects of visual data mining: (i) tasks; (ii) visual representation and (iii) the process.

Visual data plays a major role in knowledge extraction. Educational experts suggest visual tools as better process for knowledge acquisition and better communication even for a human expert. The word Visual data mining should be redefined for better use as set of algorithms models which extract data from images collected using visual aids for acquiring required knowledge for taking decisions. The term visual aid refers to human eye, a Video or digital Camera or a satellite. A human Expert extracts Knowledge for taking decisions through his sight applying natural Visual data and image mining algorithms functioning in this mind. Applying such concepts enables a computer based system to take decisions through extracting knowledge from images collected from digital device.

6.1 Visual Data Mining Techniques

- ✓ Computer Graphics
- ✓ Multimedia Systems
- ✓ Human Computer Interface
- ✓ High performance Computing
- ✓ Pattern Recognitions

6.2 Visual Data Mining Tools

- **InfoZoom**, an easy to use visual data analysis tool that works with multiple databases without SQL-querying.
- **InstantAtlas™**, improves the visual communication of location-based statistical data, with reporting solutions that combine statistics and map data.
- **IRIS Explorer**, advanced visual programming environment for 3-D data visualization, animation and manipulation
- **JMP** offers significant visualization and data mining capabilities along with classical statistical analyses.
- **K.wiz** provides a range of Java visualization components including heat maps, decision trees, 3D scatter plots and association rules.

7. Web Mining

Web mining is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining. Web mining is the application of data mining techniques to large web data repositories. The Goals of Web Mining also include the improvement of web design and structure, and generation of dynamic recommendations. Web mining is to apply data mining techniques to extract and uncover knowledge from web documents and services.

7.1 Applications for Web-Based Organizations

- Electronic Commerce
- Effective and Efficient Web Presence
- Intra-Organizational Applications

7.2 Web mining tools

- ✓ **Screen-scaper**: Screen-scraping is a tool for extracting/mining information from web sites. It can be used for searching a database, SQL server or SQL database, which interfaces with the software, to achieve the content mining requirements. The programming languages like Java, .NET, PHP, Visual Basic and Active Server Pages (ASP) can also be used to access screen scraper.
- ✓ **Web Info Extractor (WIE)**: This is a tool for data mining, extracting Web content, and Web content analysis. WIE can extract structured or unstructured data from Web page, reform into local file or save to database, place into Web server.
- ✓ **ANGOSS Knowledge Web Miner**: combines ANGOSS Knowledge STUDIO with proprietary algorithms for click stream analysis, Acxiom Data Network, and interfaces to web log reporting tools.

8. Conclusion

This paper presents a survey on various mining and tools which are used. These days mining techniques to be applied in many real time applications and industries for prevention, control and monitor all regular activities because huge data to be collected per data, it can represents in any form like image, text, numerical, audio, video, etc... Based on data mining techniques to be varied but the purpose of data mining is to get knowledge data.

REFERENCES

- [1]. Kamde, Pravin M., and Dr Algur. "A survey on web multimedia mining." *arXiv preprint arXiv:1109.1145* (2011).
- [2]. Simoff, Simeon J., Michael H. Böhlen, and Arturas Mazeika. "Visual data mining: An introduction and overview." *Visual Data Mining*. Springer Berlin Heidelberg, 2008. 1-12.
- [3]. AlZoubi, Wael Ahmad, Khairuddin Omar, and Azuraliza Abu Bakar. "An efficient mining of transactional data using graph-based technique." *Data Mining and Optimization (DMO), 2011 3rd Conference on*. IEEE, 2011.
- [4]. AlZoubi, Wael Ahmad. "Mining Medical Databases Using Graph based Association Rules."
- [5]. Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." *Kdd*. Vol. 96. 1996.
- [6]. Han, Jiawei, and K. Koperski. "Spatial Data Mining: Progress and Challenges Survey Paper." *School of Computer Science Simon Fraser University Burnaby, BC Canada V5A 1S6* (1996).
- [7]. Simoff, Simeon J., Michael H. Böhlen, and Arturas Mazeika. "Visual data mining: An introduction and overview." *Visual Data Mining*. Springer Berlin Heidelberg, 2008. 1-12.
- [8]. Pan, Jia-Yu. *Advanced Tools for Video and Multimedia Mining*. Diss. Columbia University, 2006.

Websites

- [9]. <http://image-mining.winsite.com>
- [10]. http://en.wikipedia.org/wiki/List_of_text_mining_software.
- [11]. <http://www.kdnuggets.com/software/text.html>
- [12]. <http://www.spatialdatamining.org/software>
- [13]. <http://en.wikipedia.org>
- [14]. <http://www.kdnuggets.com/2013/09/mikut-data-mining-tools-big-list-update.html>
- [15]. http://docs.oracle.com/cd/B14117_01/appdev.101/b10826/sdo_sam_concepts.htm
- [16]. <http://www.screen-scraper.com>.