

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 5, Issue. 1, January 2016, pg.233 – 240

A Novel Survey on Different Mining Tools

N.Deepika

Dept of CSE, New Horizon College Of Engineering, India

deepikajvijay@gmail.com

Abstract- From past many years Mining has been a wide research area for many researchers because of huge amount of data and information available in databases, and information repositories and Internet . With such amount of data, there is a need for powerful techniques and tools which can handle the data in better way. This paper concentrates on number of some of the free/commercial available data mining and knowledge discovery tools and software packages used in Different Mining technologies including Data Mining, Text Mining, Web Mining, Spatial Mining etc., Various Data Mining functionalities include Data cleaning, Data transformation, Data Reduction, Association analysis, Classification, Clustering ,Outlier analysis etc., For doing all these functions there are many different Data mining Tools available. Our paper concentrates to show the wide range of existing software tools with their concise features. All most all the tools are very useful in their respective task and quite easy to adopt for daily work. All Tools had some strengths and weaknesses in comparison to each other. This can be helpful for the researchers for choosing in their research. This paper presents a summary of different mining tools.

Keywords - Data Mining Tools, Weka, Rapid Miner, Orange, Text Mining tools, Web usage Mining Tools.

I. INTRODUCTION

The volatile expansion of Data ranging from several terabytes to petabytes from several data repositories such as programmed data collection tools, database systems, Web, computerized society, Business: Web, e-commerce, transactions, stocks,etc.,,Science: Remote sensing, bioinformatics, scientific simulation, etc.,,Society and everyone: news, digital cameras, YouTube .

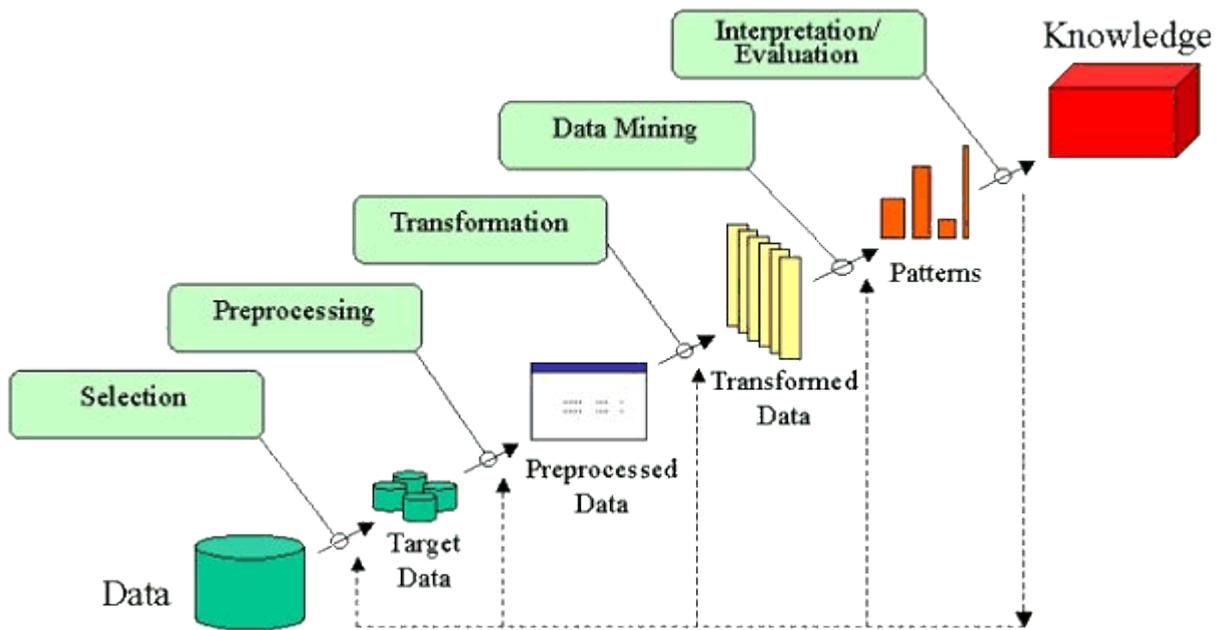


Fig 1 Knowledge Discovery Process

Because of this much rich data we are drowning in data, but famished for knowledge! Data mining (knowledge discovery from data) is the process of Extracting interesting (non-trivial, implicit, previously unfamiliar and potentially valuable) patterns or knowledge from huge amount of data. other names of Mining are Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc. The process of extracting knowledge from databases is shown in the above Fig 1[16]. Hand, Mannila and Smyth [1] define the term data mining as "Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner". Feldmans and Sangers [2] definition of text mining: Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. In a manner analogous to data mining, text mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns. In the case of text mining, nevertheless, the data sources are text collections, and appealing patterns are found not among courteous database records but in the unstructured textual data in the documents in these collections. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing activities at a Web site. It includes mining web server data ,application server data and application level data. Clustering is a method which groups the objects into groups called clusters. This is done by classifying the objects. The difference between clustering and categorization is, according to Feldman and Sanger [2], that In classification problems we are provided with a collection of preclassified training examples, and the task of the system is to learn the description of classes in order to be able to classify a new unlabeled object. In the case of clustering, the problem is to cluster the given unlabeled collection into meaningful clusters without any former information.

II. DATA MINING APPLICATIONS

Any labels connected with objects are obtained solely from the data. *Data mining applications* are constantly budding in various industries to provide more hidden knowledge that increases business competence and grows businesses. Data Mining is used in a variety of fields as below,

A. *Data Mining Applications in Sales/Marketing*

Data mining enables businesses to understand the concealed patterns inside chronological purchasing transaction data, thus helping in scheduling and initiating new marketing campaigns in rapid and cost effective way. The following illustrates several data mining applications in auction and marketing- Data mining is used for market basket analysis to present information on what product combinations were purchased together, when they were bought and in what sequence. This information helps businesses promote their most profitable products and maximize the earnings. In addition, it encourages customers to purchase related products that they may have been missed or overlooked, Retail companies uses data mining to recognize customer's manners buying patterns.

B. *Data Mining Applications in Banking / Finance*

Several data mining techniques e.g., circulated data mining have been researched, modelled and developed to help credit card fraud detection, Data mining is used to recognize clients loyalty by analyzing the data of customer's purchasing activities such as the data of frequency of purchase in a period of time, total pecuniary value of all purchases and when was the last purchase. After analyzing those dimensions, the relative measure is generated for each customer. The higher of the score, the more relative loyal the customer is, To help bank to retain credit card customers, data mining is applied. By analyzing the past data, data mining can help banks forecast customers that likely to alter their credit card association so they can plan and instigate different special offers to preserve those customers, Credit card expenditure by customer groups can be identified by using data mining, The concealed correlation's between diverse financial indicators can be exposed by using data mining, From past market data, data mining enables to identify stock trading rules.

C. *Data Mining Applications in Health Care and Insurance*

The intensification of the insurance industry completely depends on the ability of converting data into the knowledge, information or cleverness about customers, competitors and its markets. Data mining is applied in insurance industry lately but brought fabulous competitive compensation to the companies who have implemented it productively. The data mining applications in insurance industry are-Data mining is applied in claims investigation such as identifying which medical measures are claimed jointly, data mining enables to forecasts which customers will potentially purchase new policies, Data mining allows insurance companies to detect dangerous customers, behaviour patterns, Data mining helps detect deceptive behaviour.

D. *Data Mining Applications in Transportation*

Data mining helps determine the distribution schedules among warehouses and outlets and analyze loading patterns.

E. *Data Mining Applications in Medicine*

Data mining enables us to distinguish patient activities to see inward bound office visits, helps to identify the patterns of successful medical therapies for different illnesses.

F. Financial Data Analysis

The financial data in banking and financial industry is usually consistent and of elevated quality which facilitates regular data analysis and data mining. Some of the typical cases are planning and building of data warehouses for multidimensional data analysis and data mining, Loan payment forecast and customer credit policy examination, organization and clustering of patrons for embattled marketing, Detection of money laundering and other financial crimes.

G. Retail Industry

Data Mining has its great application in Retail Industry because it collects large amount of data from sales, consumer purchasing record, goods shipping, expenditure and services. It is normal that the amount of data collected will continue to increase hastily because of the increasing easiness, availability and reputation of the web. Data mining in retail industry helps in identifying customer buying patterns and trends that escort to enhanced quality of customer service and good quality customer preservation and contentment. The list of examples of data mining in the retail industry are design and creation of data warehouses based on the benefits of data mining, Multidimensional examination of sales, clients, merchandise, instant and region, Analysis of effectiveness of sales campaigns, Customer preservation, Product suggestion and cross-referencing of objects.

H. Telecommunication Industry

Nowadays telecommunication industry is one of the largely rising industries providing a variety of services such as pager, cellular phone, messenger, images, e-mail, web data transmission, etc. Due to the growth of new computer and communication technologies, the telecommunication industry is quickly expanding. This is the motive 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 of examples for which data mining improves telecommunication services –Multidimensional Analysis of Telecommunication data, deceptive pattern analysis, Identification of strange patterns, Multidimensional association and sequential patterns analysis, Mobile Telecommunication services, Use of visualization tools in telecommunication data analysis.

I. Biological Data Analysis

In current times, we have seen a marvellous expansion in the field of natural science such as genomics, proteomics, functional Genomics and biomedical research. Biological data mining is a very important part of Bioinformatics. Following are the aspects in which data mining contributes 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.

J. 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. A large amount of data sets is being generated because of the fast numerical simulations in various fields such as climate and ecosystem modelling, chemical engineering, fluid dynamics, etc. Following

are the applications of data mining in the field of Scientific Applications such as Data Warehouses and data preprocessing, Graph-based mining, Visualization and domain specific knowledge.

K. Intrusion Detection

Intrusion refers to any kind of action that threatens integrity, confidentiality, or the availability of network resources. In this world of connectivity, security has become the major issue. With increased usage of internet and availability of the 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.

III. DATA MINING TOOLS

TABLE 1
SOFTWARE SUITES FOR DATA MINING, ANALYTICS, AND KNOWLEDGE DISCOVERY

TOOL	FUNCTION
Advanced Miner	Provides a wide range of tools for data transformations, Data Mining models, data analysis and reporting
Apteryx	Offering Strategic Analytics platform, including a free Project Edition version
Angoss Knowledge Studio	A comprehensive suite of data mining and predictive modelling tools; interoperability with SAS and other major statistical tools
BayesiaLab	A complete and powerful data mining tool based on Bayesian networks, including data preparation, missing values imputation, data and variables clustering, unsupervised and supervised learning
CMSR Data Miner	Built for business data with database focus, incorporating rule-engine, neural network, neural clustering (SOM), decision tree, hotspot drill-down, cross table deviation analysis, cross-sell analysis, visualization/charts, and more
DBMiner 2.0 (Enterprise)	Powerful and affordable tool to mine large databases; uses Microsoft SQL Server 7.0 Plato
Ghost Miner	Complete data mining suite, including k-nearest neighbours, neural nets, decision tree, neurofuzzy, SVM, PCA, clustering, and visualization
Weka	Collection of machine learning algorithms for solving real-world data mining problems. It is written in java and runs on almost any platform
KEEL	Includes knowledge extraction algorithms, pre-processing techniques, evolutionary rule learning, genetic fuzzy systems, and more
KNIME	Extensible open source data mining platform implementing the data pipelining paradigm (based on eclipse)
Rapid Miner	A leading open-source system for knowledge discovery and data mining
TANAGRA	Offers a GUI interface and methods for data access, statistics, feature selection, classification, clustering, visualization, association and more
Vowpal Wabbit (Fast Learning)	Able to learn from tera-feature datasets with parallel learning, supports classification, regression, and more
CLAVIN	Software package for document geotagging and geoparsing that employs context-based geographic entity resolution
ADAMS: Advanced Data mining And Machine learning System	Free version of Alteryx, delivers the data blending, analytics, and sharing capabilities of Alteryx with just enough allowed runs of your workflow to solve one business problem or to complete one project

IV. TEXT MINING TOOLS

TABLE 2
SOFTWARE SUITES FOR TEXT MINING / TEXT ANALYTICS SOFTWARE

TOOL	FUNCTION
Aika	For mining frequent patterns within text, using ideas from neural nets and grammar induction
Data Science Toolkit	Includes geo, text, NLP, and sentiment analysis tools
Datumbox	A free API and many functions for Sentiment Analysis, Language Detection, Topic Classification and easily building intelligent apps.
IKANOW Infinite open source Community Edition,	A scalable framework for collecting, storing, processing, retrieving, analyzing, and visualizing unstructured documents and structured records
Open Calais	An open-source toolkit for including semantic functionality within your blog, content management system, website or application
Ranks.nl	Keyword analysis and webmaster tools.
Vivisimo/Clusty	Web search and text clustering engine
TXM - Unicode, XML, TEI text/corpus analysis platform,	Including graphical client, based on the CQP search engine and R environment
Alcester	A software for the automatic analysis of textual data (open questions, literature, articles, etc.)
Clara bridge	Text mining software providing end-to-end solution for customer experience professionals wishing to transform customer feedback for marketing, service and product improvements.
DataRPM	Offering Natural Language Question Answering and Automatic Data Modelling.
Eaagle text mining software,	Enables you to rapidly analyze large volumes of unstructured text, create reports and easily communicate your findings
DataRPM,	Offering Natural Language Question Answering and Automatic Data Modelling.
Active Point	Offering natural language processing and smart online catalogues, based contextual search and Active Point's TX5(TM) Discovery Engine.

V. WEB USAGE MINING TOOLS

TABLE 3
SOFTWARE SUITES FOR WEB MINING / WEB USAGE MINING SOFTWARE

TOOL	FUNCTION
SpeedTracer	A World Wide Web usage mining and analysis tool, was developed to understand user surfing behaviour by exploring the Web server log files with data mining techniques.
A1WebStats	See individual details about each website visitor, including company names, keywords, referrers, and a lot more. Free trial for 30 days.
ANGOSS KnowledgeWebMiner	Combines ANGOSS Knowledge STUDIO with proprietary algorithms for clickstream analysis, Acxiom Data Network, and interfaces to web log reporting tools.
Amadea Web Mining,	Includes multiple transformations, reports, and parametric and modular marketing indicators for an effective CRM.
AlterWind Log Analyzer Professional	Website statistics package for professional webmasters, with standard log analyzer features and unique features for SEO (Search Engine Optimization) and website promotion.
123LogAnalyzer	Simple to use, high-speed processing, low disk space requirements, filtering, and built-in IP mapping
XAffinity(TM	For identifying affinities or patterns in transaction and click stream data
Web Trends	A suite for Data Mining of web traffic information
SPSS Web Mining for Clementine	Enables you to extract Web events, including online campaign results, and use this online behaviour in Clementine's predictive modelling environment.
The Data Miner	A tool for automating web data extraction and manipulation
Surf Pattern Visual Analyzer	Web navigation visual link analysis tools showing web page viewing patterns of a web-site
Affinium Net Insight	Unica enterprise-class web analytics.
Click Tracks	Displays visitor patterns directly on the pages of your website.

Nihuo Web Log Analyzer	Provides a comprehensive analysis of the "who, what, when, where and how" of customers visited your web site.
SAS Web hound	Analyzes Web site traffic to answer questions like: who is visiting; how long do they stay? What are they looking at?
Download Analyzer	Can track visitors, hits, downloads, referring sites and search phrases, and provide traffic analysis data for web promotion and search engine optimization.
Megaputer Web Analyst,	Integrates the data and text mining capabilities of Megaputer's analytical software directly into your website.
Conversion Track from Ants soft	Web log analysis and reports on visitor conversion ratios.

VI. CONCLUSION

Data mining is the process of drawing out valuable patterns and associations from data sources, such as databases, texts, the web. Using data mining to recognize and extrapolate information and in turn can reduce the probability of deception, progress review reactions to prospective business changes, and make sure that risks are managed in an additional well-timed and practical manner. Auditors also can use data mining kit to model "what-if" situations and display actual and feasible things to organization, such as combining real-world and company information to illustrate the special effects of a safety infringe and the impact of trailing a valuable patron.

REFERENCES

- [1] Hand, D., Mannila, H., Smyth, P. 2001. Principles of Data Mining. Massachusetts Institute of Technology.
- [2] Feldman, R., Sanger, J. 2007. The Text Mining Handbook. Cambridge University Press.
- [3] J. Han, M. Kamber: Data Mining. Concepts and Techniques , Second Edition, Elsevier, 2006;
- [4] M. Andronie, M. ANDRONIE: Analiza datelor stocate in depozite mari de date , Sesiunea de comunicari stiintifice a cadrelor didactice din facultățile economice ale Universității Spiru Haret, Bucure ști, 2008, ISBN 978-973-163- 230-8;
- [5] G. Piatetsky-Shapiro, U. M. Fayyad, and P. Smyth. From data mining to knowledge discovery: An overview. In U.M. Fayyad, et al. (eds.), Advances in Knowledge Discovery and Data Mining, 1-35. AAI/MIT Press, 1996.
- [6] Mark Hall, Eibe Frank, G. Holmes, B. Pfahringer, P. Reutemann, IH Witten, The WEKA data mining software: An update ACM SIGKDD Explorations, Newsletter, Pages 10-18, volume 11 issue 1, June 2009.
- [7] <http://rapid-i.com/> [5] <http://eric.univ-lyon2.fr/~ricco/tanagra/>
- [8] DBMiner: A System for Data Mining in Relational Databases and Data Warehouses, Data Mining Research Group, Intelligent Database Systems Research Laboratory School of Computing Science, Simon Fraser University, British Columbia, Canada, <http://db.cs.sfu.ca/DBMiner>.
- [9] www.uea.ac.uk/polopoly_fs/1.3589!introductionkdd.pdf
- [10] <http://orange.biolab.si/>
- [11] Laura Ruotsalainen, Data Mining Tools for Technology and Competitive Intelligence, espoo 2008 vtt research notes 2451
- [12] Mihai ANDRONIE ,, Daniel CRI ȘAN , . Commercially Available Data Mining Tools used in the Economic Environment Database Systems Journal vol. I, no. 2/2010.
- [13] Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam, A Study of Data Mining Tools in Knowledge Discovery Process, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012
- [14] <http://www.rdatamining.com/resources/tools>
- [15] <http://www.predictiveanalyticstoday.com/top-free-data-mining-software/>
- [16] http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html



N.Deepika, Sr.Asst.Professor having 14 years of experience in Academics has pursued her M.Tech from JNTU, Hyderabad. She is currently working In NHCE, Dept of CSE, Bangalore. She has guided many UG & PG students for their Projects. Her Research areas include Clustering techniques, Data Mining, Web Mining and Big Data Analysis.