A Survey on Data Mining Approaches

Amirmahdi Mohammadighavam
Department of Computer Science,
Shahid Bahonar University,
Kerman, Iran

Neda Rajabpour
Department of Computer Science,
Shahid Bahonar University,
Kerman, Iran

Ali Naserasadi
Industrial and Mining Faculty,
Shahid Bahonar University,
Kerman, Iran

ABSTRACT
In the new world, with advances in modern industries and businesses and increase in the growing need for data with reference to the knowledge of data processing in order to participate in a big competitive market, led all to the new techniques for mass processing of raw data on their role in data mining due to the direct use of the machine And thereby increase the speed, accuracy and quality of information is much more important than other methods. Hence in this paper the concept and meaning of the data mining techniques and models used in data mining and data mining is the major subtype web mining has been studied. Finally, we have examined and reviewed the latest achievements in data mining and new approaches that have covered.

General Terms
Data Mining

Keywords
Data Mining, Web Mining, Text Mining, Mass Processing, Machine Processing.

1. INTRODUCTION
In today's world competition is one of the most important challenges facing all organizations and industries. That is hard to find in a particular organization or industry which has no rival to him.

Between the industries and organizations could remain successful and lasting good that know their customers need and take steps according to their needs. On the other hand, in today's world that the era of data explosion , all organizations and industries have so much data that they have even no processing of the data because of their lack of storage areas are replaced with new data and they are replaced with new data storage and the process is repeated continuously.

High speed and full use of machines capable of processing speed and directional data in different fields and also the need to make assumptions and registration in search of growth in IT hardware led to the recent addition of oral and manufacturing and service industries even academic bodies and research in different fields such as meteorology and geology, and medicine and many other sciences has made data mining as a technique for extracting knowledge from data.

Despite all the advances in science and technology, unfortunately have always been involved in some data mining machine that compare with some methods for making such as (OLAP) and such titles or even worse to read the titles with them. Here we have tried to talk about the concept and nature of the similarities and differences in data mining and data mining needs of the other measure is to clarify the concept of data mining then examines the models and data mining techniques and tools necessary to have covered and then talk about one of the main branches, means web mining.

In continue the most important part of the discussion on new applications as well as the science of using data mining to discover the letters or use data mining to detect fraud and new approaches such as the ability to use the science of data mining in geological materials is described.

2. DATA MINING PROCESS AND METHODOLOGY
Data mining is a process to extract reliable information and keep track of the massive data set. Data mining discovers patterns and trends in the data. The model and methods can be together and be defined as a search model. Provide a data mining model and the dynamic and repeatable process .Exploring the creation model is a part of a larger process that is defined in the receiver to run the model in the workplace. [1]

Data mining is: Discovering the methods and patterns in large databases to guide decisions about future activities. It is expected that data mining tools to get the model with minimal input from the user to recognize. The model presented can be useful to understand the unexpected and provide an analysis of data followed by other tools to put decision-making are examined and it ultimately leads to strategic decisions and business intelligence. The simplest word for knowledge extraction and exploration of volume data is very high and the more appropriate name for this term is "Exploring the knowledge of database". A database is knowledge of discovery process. This process includes the preparation and interpretation of results. The Gartner Group consultancy for data mining "the process of extracting meaningful new correlation pattern by an inspection methods and exploring the large volumes of data are stored in the data warehouse. [2, 3, 4, 5]

3. DESPITE THE NEDD FOR DATA MINIG
Rapid and dramatic growth of data collected and stored in their databases, many of the human ability to perceive and understand it is not possible without powerful tools. Gather data in the database data to the tomb has become. The outcome of important decisions based on the rich information stored in databases and decision-makers have the tools to extract knowledge hidden in databases were not huge. The amount of data doubles every 5 years are available and capable organization that is able to manage less than 7 percent of its data. Data mining is a major concern because the information industry in recent years and is located in a high

volume of data and work across a wide and decision makers in information-rich resource utilization are unable to make real decisions and the availability of data in commercial trade are lack of knowledge. [3, 6, 7]

It is clear that large amounts of data are aggregate. However, what these data can be achieved. In early 1984 we've been drowning in data while we were hungry for knowledge. In fact, in most areas are awash in data and the problem is that analysts do not have enough skill and experience necessary to have knowledge in the translation. [6]

Significant growth in data mining and knowledge extraction from a confluence of factors has been the strength of various factors:

1. Exponential growth of aggregate data to be produced and the availability of large volumes of data.
2. Data storage and analysis of a data warehouse and analysis database and on-line analytical processing technology development.
3. Increase the amount of data and internet access.
4. The dramatic growth of computing power and storage space.
5. Development of data mining software products and software that are readily available in commercial data mining and data mining applications, user interfaces are standard.

Rising competition in the global economy and the stock market to manage the client's interest [1, 4, 7].

4. AMONG THE DATA MINING OF SCIENCES

The roots of data mining in three of the track are read. This is the classic family hit. There will be no data mining technology that is so often the basis of statistics that are based on data mining. Statistics of classical concepts such as regression analysis and distribution of standard deviation and variance, and cluster analysis and the confidence intervals for all of these cases were studied and the relationship between the data is. Classical statistical analysis techniques, data mining plays a key role. The second belongs to a family of data mining, artificial intelligence is. AI-based approach is innovative and with opposition figures and tried to think of such a process to resolve the question of statistical apply. Because this approach requires high computing power is not practical until the early 1980s. [8]

5. THE SYSTEM ARCHITECURE BASED ON DATA MINING

1. Database, storage and analytical data, or other repository of information that includes a set of databases, spreadsheets, and analytical data storage and refining techniques and on the accumulation of data is performed.
2. Server database or data warehouse and analytical data associated with the request that fetches the data mining user.
3. Bank of knowledge: domain knowledge to guide the research and evaluation results are used in interesting patterns.
4. Data mining engines**:** the main components of data mining systems and includes a set of functions for data mining tasks are.
5. Patterns**:** knowledge of the patterns presented in the form and accuracy by the functions they will be evaluated.
6. User Interface: the communication between the user and the system is data mining and visualization tool for exploring patterns of different forms. [3**]**

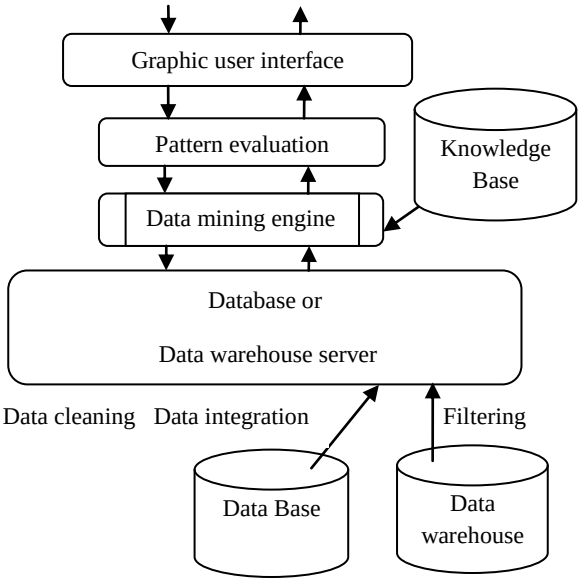

Fig 1: architecture of a typical data mining system [3]

6. WEB MINING

Web mining , significant and robust data model used by the contents of Web pages, the information available on the Web, Web connections and e-commerce is using data mining techniques , the practical result of these techniques to help users extract information better, Designed to improve and expand e-commerce is better.

6.1 Classification of Web Mining

The field of Web mining and processing, is divided into several categories. Most scholars and researchers, web mining into three sub-categories of web usage mining, web content mining, web structure mining are divided.

6.1.1 Web content mining

Discover and extract useful information from the content, data, text and web are in this category. In the past, different types of Internet services and resources such as Usenet, ftp, Gopher, was formed today; more data are available on the Web. Over the past years, information on the Web have grown very rapidly and we see that today's digital library, the Web has expanded rapidly and many companies have adopted e-commerce and its services, and many companies large and small, that their local database and available only to put themselves but now through the web, we can find a huge database of companies anywhere in the world, to have access.

Thus we see the web content mining, and a wide range of information to be included. Over the past years and have very fast growth of the Web and digital libraries and on the web today, we have expanded rapidly and many companies Have their own e-commerce and Many companies large and small that their local databases and their availability would. Now we can through the web to huge databases these companies have access from anywhere in the world. Thus we see web content mining And a wide range of information to be included. [9]

6.1.2 Web usage mining

Web usage mining is one of the following directories on the web mining techniques that can interact with the Web user behavior to predict, Focuses on data obtained with the latter are derived from the interactive website, and has worked. Organs are usually very large volume of data in their daily practice by a web server in the server access log file to collect. The analysis of this data can help these organizations to become lifetime customers and sales strategies with respect to the production of crops, the end user on a website, URLs that have access to the web and other information from this class of server log files have been extracted. [10]

6.1.3 Web structure mining

These sub groups of web mining trying to explore and investigate a model based on links in web. This model is based on topology hyperlink and can be grouped into Web pages and links between pages on a site and between sites as well as to extract.

Web structure mining, structure hyperlink read the Web and web pages can be grouped and then discovers the connection between websites. Here are the algorithms used, page rank and hyperlink induced topic. Thus the subset web structure mining can be summarized as:

1. The study of hyperlink
2. Category Pages of web page
3. Summary of information**:** Connections and similarities between web pages [10].

7. THE PROCESS OF DATA MINING IN CRM

Data mining is one of the elements of customer relationship management (CRM) and can help move the company towards customer orientation. Data mining in customer relationship management process is as follows: Raw data from various sources are collected and the extraction and translation, and process management into the data warehouse are callable. The model consists of four layers is found:

1. Questions such as the customer's business.
2. Applications such as scoring, predictive.
3. Methods such as time series, classification.
4. Algorithms

In this section, data mining methods with special application to answer for your businesses to reach the mind, the algorithms are derived and algorithms for the construction of the model used.

In the analysis model, the patterns are converted into a useful and usable knowledge and the improvement of their models that are efficient in an operational system will be used. [11]

8. MONEY LAUNDERIN DETECTION

With increasing global trade growth and expanding international market, through the exchange of money and currency will rise. In this way become the property of the non-traditional ways to make money with it and clean background in terms of money laundering to say that every day can increase.

Other bands that intercept the mafia and black market dealers of the traditional way is impossible, these factors along with the growing new technologies in the financial thinkers to explore new ways of corruption, including money laundering, which in turn the data mining technology is a special place for the coordination of this issue as this article explains the overview. [11]

8.1 Examples of Money Laundering:

First case:

Crime is a business wants a million dollars from import of goods to money laundering. First is a local importer, exporter with a foreign (maybe himself) to collusion. Then perform the following tasks:

1. The foreign exporter pays one tenth cents per ten thousand to buy razors. (Totaling $ 1,000).
2. Tens of thousands of foreign exporters sell the blade to a local exporter for $ 100 (Totaling $ 1 million).
3. Local importer of ten thousand dollars to get the blade to the actual price and pays a million dollars to the foreign exporter.
4. Local importer to a foreign country with a million dollars to the cost (is washed).

The addition of these bills in the importation takes place.

Second case:

Suppose a business wants a million dollars through exports crime to launder money. First is a local issue, to conspiring with a foreign importer. Then perform the following tasks:

1. A local misdemeanor, with a million dollars, 200 hours for each five thousands dollar price of gold luxury buys with cash.(totaling one million dollar)
2. The local exporter, an importer of foreign gold 200 hours to the selling price of 5$ each. (totaling one thousand dollars)
3. The foreign importer, received 200 hours of gold and a thousand dollars to account for local exporters.
4. A foreign issuer, the market price of gold in each sells five thousand dollars.(totaling one million dollars)

Results: The local issuer of a million dollars to a foreign country with a $ 1,000 transfer fee (washed) is.

The above procedure with a low score on the export of goods takes place.

The use of trade to transfer black money from one country to another is one of the old approaches to escape from public accountability. This added to the invoice billing to reduce the import or export carried out. But unlike both of the above

possibility, economic and intelligence agencies are doing to detect money laundering through financial institution and accounting data from international trades.

The high import bills on Tuesday may be the opposite:
1.the customs.
2. Evasion.
3. Money laundering.

This money may be available to an organization like al Qaeda. Money laundering operations and the way any country will be the best way to study the environment. for example, fewer controls on exports to other countries. [12, 13]

Table1. Money moved from the U.S. to Al Qaeda watch list countries. [13]

Country	Dollar amount moved
Malaysia	$2,220,978,718
Indonesia	$ 564,597,632
Saudi Arabia	$ 486,669,284
united Arab Emirates	$ 232,737,819
Egypt	$ 148,085,489
Total-Top 5 countries	$3,653,068,906
Other Countries-20	$ 619,142,176
TOTAL	$4,272,211,082

8.2 Related Data Mining Methods

data import and export goods in the united states department of commerce and census data show that commercial goods can be used with data mining techniques to find clues. with an estimated price of goods exported and imported goods at a certain time can be examined and this is good for 16,390 import and export in 2001 and 8568 for 230 countries that had a business engagement with the united states was. All imports and exports were recorded compared with the high and low. Dollar amounts and number of suspected cases was the accumulation of any country. The total amount money transferred out of the United States in 2001 was 156 billion$ and this was monitored by means of all suspected cases [12, 13].

9. TYPICAL APPLICATIONS IN BUSINESS

The sales forecasting process, trade, commerce and trade guarantee, quality control, personal issues, including commercial banking and the non-oil industry, science, fire prevention, detection, chemical structure, crime detection and diagnosis, is efficiency our main focus now is on the organization of data warehouses most of the data analysis process and there , the data is complete. summarize the data stored about them and do more research and take them to the next and subsequent analysis takes these comments are summarized in the truth of the information ,are extracted. For more research is done to develop other ideas in this section, several examples of data mining in the real world around us. [1, 4]

9.1 Finance & Banking

U.S. banks already have demographic theories to explain the functions and financial assets watch selected groups of customers, have prepared and presented. With numerous studies, the 800 GB of data, on average within 30 seconds, is stored security services of data mining is used to determine how the financial markets to destabilize the security of businesses, react. For example, relationships between data exchanges between the Japanese yen and the market is how the government? Royal Bank of Canada's business is based in Toronto for the maintenance of currency, trade patterns, the relative mix of products sold and products to design and implement marketing strategies and performance criteria, is defined. Data stored in a fixed surface and a surface suitable for research in particular places. [14]

10. EXTRACTION OF GEOLOGICAL INFORMATION

Of field-scale measurements to simulate the weather (climate) and the remote sensing (GIS) and the large and growing volume of earth science data, hard analysis, visualization, simulation and interpretation will increase.

Data mining (information extraction), information theoretic (information theory), machine learning technique (techniques of machine learning), cluster analysis, singular value decomposition (decomposition of single-valued), block entropy (block entropy) “block entropy = the measured thermodynamic) phase-space reconstruction (phase-space modeling), artificial neural network (artificial neural networks) to solve problems such as clustering, the feature extraction, tracking changes, compare and validate models to data models are used. The size and complexity of earth science data analysis tools and much more limited capacity goes beyond desktop computers. New tools to analyze and visualize scaled on the batch and parallel computers and supercomputers are implemented, the data in this size and scale can be done.

10.1 The Earth Science Data

Combination, integrating and incorporation of data from earth system science create new opportunities for scientific discoveries that are known to be create d. Information (data centric science) to offset those involved in information, create a worse situation enters the environment, caused by the use of land users and decision makers in ground increases. [3, 15]

10.2 Information Extraction Approaches

Wide variety of data mining, machine learning techniques and theoretical data (information theoretic techniques) also led to their use in the body of scientific data to earth (earth science data) grow has been demonstrated that cluster analysis (cluster analysis for clustering (segmentation), extraction of features, a network analysis (change detection), an internal comparison) and compared the data model earth sciences in the number of users is effective. Using artificial neural network foe data modeling and classification of earth sciences (earth science) rises. In a report, use ANNS, neural networks for the purposes of refining the structural model images to create a full-wave tomography “ lesions detected by light-writing“ are used to represent data. IANN ANNS and has led to the use of classified information from the wave forms produced by seismographs and pain stacking process large volumes of production of high quality images of the layers decreases.[15]

11. FRAUD AND SPAM EMAILS DETECTION

Still a lot of problems with spam for the world have in addition to their mail fraud and identity theft and cyber security in the world challenge and the discovery of the hoax letters in an electronic world will increase security and prevent identity theft. Many efforts to prevent the spread of

spam are based on data mining algorithms .Due to the nature of his new trick, but the letters are lower case. A new method of detecting deception letters of the algorithms based on the content of the letter is a hoax. One of the ways to detect spam and hoax letters in search of information data such as address, number of recipients and is the volume of letters. The new trick is the nature of letters that are under consideration. Development cheat sheets, are seeking new ways to deal with them. [16, 17]

Letters of deceit has shown that most of these programs varied from the use of certain words and phrases New methods and new separation of other deceives to other letters of the letter has been data mining algorithms. The method is based on the content of this letter using text classification algorithms to classify e-mails to trick and deceive the two classes. One of the most important branches of data mining for text classification is based on the content of texts. So far, many algorithms have been proposed to classify texts. [18, 19]

There are several ways to detect spam. Usually these letters with the help of network information such as address, number of recipients, and the volume of spam that are available in the draft, are detected .However, in the year recent data mining techniques to help prevent spam from being sent. But looks can be deceiving letters in the pattern of data mining algorithms to discover some of them can be prevented. Using data mining algorithms and classification of electronic texts can be categorized based on the text. [17, 18]

11.1 E-World Security Challenges

Cyber Host & Domain with the electronic world, these areas has witnessed the development of specific security challenges. One of the most important challenges the world of cyber security, fraud is discussed. According to the FTC's progress in the world of cyber fraud and identity theft as much as 37%, respectively, 12% of Internet auction and the offer for foreign currency rate is 8%.

11.2 Data Mining Algorithms for Detecting Deception Letters

The issue before us is in fact a binary classification problem. The goal is to design a filter for messages, so that it can detect patterns of fraud in the hoax letters. Filter the mail fraud and deception with the help of two classes of messages are taught. Then the filter can be placed in the electronic message server and a real-time with a server and a real-time e-mail is received to review. This new sequence is the classical text mining with semantic learning in this section we classify the texts of the three algorithms for a few close neighbors. [18, 20]

12. CONCLUSION

Given the increasing need to identify new ways for humans and undeniable progress in many sciences, including international trade, manufacturing, and other science can be of Data Mining as a window to jump on this name the board.

Nowadays, we can find many industries that they have gone out of business. Because of not knowing the market or customers desires. do we name of the sciences which have reached to the end line because of human disability in processing of data's which without processing in speed of data changing we cant expect to progress in these aspects in future, for example geology, seismography and weathering are the best samples of these sciences.

Of course, by excellent processing in artificial intelligence and machine processing and increasing in machine powers, we can imagine using this kind of processing for a good science. Of course, with considering the extra need of industries occupations to enough science for decision, we can hope to a good trade fund for this fairly new science

At the end, it should be mentioned that many above usages in this essay more make you familiar with the subject and we really didn't go through the details and our main aim has been to represent ideas in new aspects and introducing the usage type of data mining in industries and sciences.

13. REFERENCES

[1] ZhaoHui Tang, Jamie MacLennan, "Data mining with SQLSERVER2005".

[2] Hamid R. Nemati and Charmion Brathwait , Kara Harrington , "Privacy Implications of Organizational Data Mining" , 2004.

[3]Jiawei Han & Micheline Kamber, "Data Mining Concept and Techniques", 2000.

[4] DANIEL T.LAROSE, WILEY INTERSCIENCE, "Discovery Knowledge in Data, An Introduction to Data Mining.

[5] David M. & Natalie M. Stiger, Stiger University of Maine, USA "Knowledge Mining in DSS Model Analysis", 2004.

[6] Hamid R. Nemati and Christopher D.Barko – University Of North Carolina at Greensboro, USA 2004"Oraganizational Data Mining (ODM): An Introduction" (Springer).

[7] Michael J.A Berry, Gordan S. Linoff Wiley - "Data Mining Techniques for Marketing Sales and Customer Support Management.

[8] Two Crows Corporation, Introduction to Data Mining and Knowledge Discovery, 1999.

[9] Web Mining Research, Raymond Kosala, Hendrik Blockeel.

[10] Web Mining, Accomplishments & Future Directions, Jaideep Srivastava University of Minnesota, USA.

[11] THE HAND BOOK OF DATA MINING, ARIZONA STATE UNIVERSITY, 2003.

[12] Application of Cluster-Based Local Outlier Factor Algorithm in Anti-Money Laundering, School of Economics and Management Southwest Jiaotong University, Gao Zengan.

[13] Detecting Money Laundering and Terrorist Financing via Data Mining, John S. Zdanowicz.

[14] Intelligent Miner for Data, Joerg Reinschmidt, Helena Gottschalk, Hosung Kim, Damiaan Zwietering.

[15] Data Mining in Earth System Science (DMESS 2011), International Conference on Computational Science, ICCS 2011, Forrest M. Hoffman, J. Walter Larson, Richard Tran Mills.

[16] D. Cook, J. Hartnett, K. Manderson and J. Scanlan, Catching Spam before it Arrives.

[17] Pfleeger, SL & Bloom, "Canning Spam: Proposed Solutions to Unwanted Email", Security & Privacy Magazine, IEEE.

[18] Airoldi, E, Malin, B. "Data mining challenges for electronic safety.

[19] Nm Y. Yang, An evaluation of statistical approaches to text categorization.

[20] N. Littlestone and M. Warmuth , "Weighted majority algorithm.