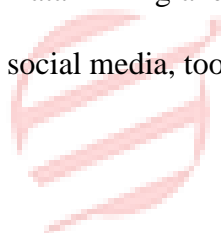DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

Data Mining and Database Technology (Web Mining, Text Mining, Sentimental Analysis for social media, tools, techniques, methods, applications etc.)

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

Abstract

This paper discusses significance of Database and data mining techniques and their multi purpose applications. IBM DB, Oracle, MySQL etc. are some example of DBMS software. KNIME, Carrot2, GATE, ELKI etc. are the examples of data mining software. Techniques of data mining and database are being used in different sectors i.e. Marketing, Manufacturing, Banking, finance, government etc. Data models are built by using historical data. Data mining is useful for quality control purpose. Financial transaction records can be maintained and analyzed with the help of Data mining. Historical data play significant role in data mining. At present, database management plays significant role in context of business operations. Law enforcement agencies maintain the crime database. Data interpretation, answering query has become very easy now. In context of datamining and database technology, there are some issues i.e. Incorrect information, Privacy, security etc. Behavior of third party is important here. With the help of data mining and database techniques, quality of decision making has improved a lot. It has become easy to study data trends. Recent technological advances have made automation of datamining easy. Extensive research can be performed using data analysis. Nowadays, many businesses are involved in data collection. Many businesses are getting competitive edge due to this. Popularity of database and datamining techniques is increasing nowadays. Technological developments in hardware and software has played significant role in this context. With the help of datamining data can be categories as well as clustered. Variation in data behavior can be detected. Regression and other statistical techniques can be applied.

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

Introduction

Data Mining and Database Technology (Web Mining, Text Mining, Sentimental Analysis for social media, tools, techniques, methods, applications etc.)

Decades belonging the importance, use and relevance of IT are in the huge increase. The study, use and demand for IT is highly developing the each and every day, like, as per the ruling concern & now, this is a necessity for men. Using the highly developed systems, like, computers & telecommunications which refer the IT is going prosperously. Belonging those, there are many successful and popular IT sectors have born and now are controlling the economy. Besides, these telephonic and normal computer base information technologies, the power of data mining and database technology which literally refers to the huge concepts of IT are also crossing their distance of milestones day by day. And on behalf of that this paper is created.

Database Technology - The DT generally refers to the DBMS (Data Base Management System), is a highly decorated computer software application, that interacts with the user, other related applications, and the data itself, to capture and analyze the data. That term DATABASE refers to a huge collection of data. The collection of tables, information, queries, schemas, reports, different views and other many objectives. This supports one company by giving information which was previously loaded, but very difficult to find out. But by a query on the successful DTs, like, MySQL, Oracle, IBM DB and many more, the finder can easily get a single word of data from cores of data. Database Technology is highly recommended for all the big business groups, for keeping and retrieving their information. Like, if you having a hotel, which's 90 rooms, are already booked among of the 115, then besides a manual

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

research, you can easily allow a vacant room to your new customer by just typing only one query, which answers you which room is vacant. So, database technology makes easy and comfortable and it is a way to develop.

Data Mining- The term data mining refers to a huge concept, which defines and proves that DM is a pillar of the IT because without DM large companies will never be able to handle their stocks through IT. It helps companies to focus on the important issues in their data warehouse by its most powerful technologies. By data mining, one can examine the pre-existing database in order to generate new information. This concept works like that, most of the valuable and international companies have a massive amount of data, now, the data mining technology can approach and implemented easily on the existing hardware and software paths to enhance and express the value of their existing resources. And the benefit is - it can be easily integrated with new systems, products as they are brought on-line. The performance also proves its potential while implanted on high-performance server/client or the parallel processing computers. The upgraded tools of data mining can easily analyze a massive database to deliver answers to such questions, like, "Which clients are most likely to respond to my next promotional mailing, and why", which makes it easier to understand. As the best part of the use of data mining is, it can answer business questions, which is, on the other hand, is too time-consuming to resolve. As an overall review, it makes the system easy; it can explore the data with all its way, it is exploring the IT, and is helping the IT & IT business world.

<div align="center">Literature Review</div>

Data Mining & Database Technology

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

Data mining is the new technology to extract out information from a given set of humongous data which was otherwise not apparent or was hidden and has the potential to direct big firms to focus on things important for their growth.

Using these techniques businesses and companies has the ability to make data and fact driven decisions. Such decisions are made by the trends in behaviors and the predictions made by these analyses. The analysis offered by data mining is automated and have outreach outside the analyzes of past given by the tools typical of decision systems.

Earlier there were queries which would require a lot of time to research on and hence were left out. Now Data mining tools have enabled the companies to get an answer to that too. Huge databases are scoured for finding the patterns which were otherwise not visible directly. It can also predict stuff that was not even expected by the experts because it was not apparent.

Currently a lot of firms are collecting data for doing such analysis. This is because data mining techniques are easy to implement even if the platform is existing, be it software or hardware. This gives an extra edge to the information sources which can then be used to create new products as and when required.

There are things that the companies need to know, like if they launch a promotion which segment of the customer will be swayed the most towards it. Such answers are achieved by the parallel processing computers.

TASKS Achieved By DATA MINING

Fayyad et.al. (1996) has shown the following six usage of data mining:

1) Detection of the important changes in the overall data with one or more variables.

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

2) Finding set of categories and clustering the data in them

3) Mapping a given function variable to a real prediction variable using regression techniques.

4) Sequencing the given data just like that of association rule

5) Modeling a system describing the inter-dependencies of variables among them

6) Analysis of the data provided and then grouping into various classes

Web Mining

When the above-described data mining techniques are used on the web pages to get relevant information the method utilized is called web mining. It is then classified on the basis of type of data used for this

a) Web Server Data: The data collected by the web server like the page access time of the IP of the user connected.

b) Application Server Data: A little modification is done on the existing commercial applications servers and then they are used to track various kinds of business events and use the data collected thereby for various purposes.

c) Application Level Data: Events can be defined in an application which do not exist currently, and data recording can be turned on for them thus creating the past of such of specially defined events

Text Mining

Text mining, or text data mining, is the method of gaining important insights from a given set of text. Methods like statistical pattern learning are used to extract information about trends and other patterns.

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

This process usually involves restructuring the input text by removing some part or adding some and insertion of the text in a database. This is followed by finding patterns in this structured data and interpreting that pattern for useful purposes.

Such analyzes also involve predictions, identifying patterns, finding distribution frequency, extracting information, analysis of clustering and associations within variables etc. The ultimate aim to make sure that the available text is turned as data which can then be easily analyzed using natural language processing (NLP).

An example of the application could be scanning a document having usual handwriting and use it for normal search indexes or extract information out of it.

Sentiment analysis

Sentiment analysis usually known as opinion mining is using text analysis for extracting information and applying language processing for getting subjective information. Sentiment analysis is used in digital marketing and social media marketing for variety of applications.

In general it gives you the overall attitude of a user for a product. This could be his/her personal judgment, the emotional state or whatever he/she wishes to express for the product.

Data Mining Tools

Some of the data mining tools as showed by Huang (2005) are:-

1)11 Ants:-used for business applications

2) ADAPA- a framework to employ and integrate

3) Data Applied- A framework for Data Analysis

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

4) Revolution R Enterprise- Based on an open source software R with many additional tools like Hadoop for big data and database coupling.

Data Mining Techniques

Data mining technique involve static machine learning, usage of databases and machine learning. Following are the two approaches for the data mining:-

1. Statistical Approach

The models based on statistics are designed using a set of pre-existing data. Theories like regression and correlation analysis or Bayesian network or clustering of data are used for data mining.

A good example can be that of traffic jam where the Bayesian Network is used. Here each node is for one state and the each edge is for interdependencies among nodes. The analysis and the data interpretation of the accidents easily show that there is high chance of accident during rush hours or poor weather which again leads to traffic.

2. Machine Learning Approach

The machine learning methods which are used most frequently for data mining include conceptual learning, inductive concept learning and decision tree induction.

An object's class is determined by a decision tree by following the path from root to leaf node, Gheware (2014).

Data Mining Applications

These days data collection has become an easy task and given the invaluable insights achieved from the data mining techniques many fields utilize data mining. Data mining

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

techniques are used in field like finance, biology, medical, marketing, banking, e-commerce etc. Following applications have been given in detail:-

1. Financial Data Analysis

The data that is available in banking and financial industry is very precise and therefore very reliable. Due to this the predictions and the interpretations from data analysis are of high quality.

Design and construction of data warehouses for multidimensional data analysis and data mining Gheware (2014). Loan payment prediction and customer credit policy analysis. Classification and clustering of customers for targeted marketing. Data trends can also be analyzed by government to detect financial crimes like that of black money or tax evasion.

2. Retail Industry

Retail industry collects data from its customers, their habits, purchasing history, sales history and goods purchase history. Now, since there is an obvious popularity of web services and the ease of usage there will be a increase in this data and its prediction. The analysis of this data gives the Retail Industry the buying trends and hence helps them give the exact promotion and services required by the customer.

3. Telecommunication Industry

The Telecommunication industry using technologies like email, web, cellular phone etc is emerging very fast and hence the popularity of data mining. This can be use to identify the patterns in calls or catch the frauds, maximize the usage of available resources and increase the service quality given to the customers.

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

4) Data Analysis on Biological Data

The humongous growth in the field of genomics has led to use of biological data mining for bioinformatics. The insights gained from bioinformatics are used to predict functions of genes and their role in genetical factors.

Challenges Faced

Although data mining gives very precise results from the data present the main problem is the presence of inaccurate and low-quality data. Other problems could also include uncertainty of the data present, none- heterogeneous data, privacy matters etc.

If the data itself the not precise the reliability of the models and results hence obtained, is very low. Methods are employed to make sure this ambiguity is removed and the data represents all the strata of the context.

Uncertain data is the problem in cases of data sets like location tracking via GPS data where the location is the mean of each field and the variances indicate the errors which might occur.

If the data is not certain that means the value of that particular item is not given as a distribution rather than a single value. Now this distribution can't directly be used for creating the data mining algorithms.

Future Trends

Sharma (2014) has emphasized to the fact that the complexities involved in the process of data mining must not be presented to its end-users. Business use cases can be designed, with tight constraints, around data mining algorithms.

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

Since there has been huge success of data mining and its applications, the field of data mining is now been seen itself as the major part of computer science and has shown interest potential for the future research.

Discussion (Advantages, disadvantages, Impacts, issues, if any, and related applications)

As per discussion is concerned, then talking about the advantages, disadvantages and also about the impacts of IT, it can't be completed in one day, and also it may take place to some arguments on some topics. IT is ruling the world and this is for sure that without giving a huge amount of success to the industry and business that would never be possible to rule in the business. The technologies of data mining and database are developing daily. Their demand, their need and after getting such IT software, their workings are beneficial for the industry the IT industry hugely.. Businesses, whatever it is small or big organizations, before building an office and building physically, they are creating and publishing a website about them. Small industries, where only 200 or 300 employees are working on the most, they are also keeping their all employee records, with the most useful details in their computer database and people are retrieving that using that with the help of data mining. Colleges, which are not so big in population, today they are using a DBMS, even at the library. So, these are the examples. And that clears the scenario about the success of database technology and data mining in the IT.

Some Examples of DT and DM- Oracle, MySQL, IBM DB these are the DBMS software. And some DM software is- Carrot2, ELKI, GATE, KNIME etc.

Now, just have a point full look at the Advantages and disadvantages and Related Applications of Data Mining & Database Technology:

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

Advantages of DM & DT:

Data Mining has a lot of advantages while using at the each and every specific industry. And it has some of its own disadvantages also. And here we will describe it by all a specific way, very briefly-

Some Data Mining and Database Technology Using Sectors –

- *Manufacturing Sectors.*

- *Finance/Banking Sectors.*

- *Marketing Sectors.*

- *Government Sectors.*

Advantages – At the marketing sectors, DM helps them for building models based, on the historical data, and provides the information that who will respond to the marketing campaigns well, like, online marketing campaign, direct mail and much more. And through the given results of DM, the marketing companies get their goal very appropriately by selling their right products to the right customers. And like that it helps the all sectors in getting their goals in their business industry. Like, in the financial sectors, DM gives them information regarding loan and credit reporting. In the manufacturing departments it helps them to explore the operational engineering data, and that will help the manufacturers to detect the faculty equipments and by that they will able to judge the optical control parameters. At the government industries, DM helps them by digging the records of financial transactions & helps to built the pattern that can easily direct any criminal activities and if there any laundering. And what makes this all things possible, it is database technology. Because without getting the previous data and information about the company, data mining will never

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

be able to give any information, because to get the best it has to dig, it has to analyze the going time and previous history. And database management is the pillar of a business or company, because if you forgot all things about your life, then can you imagine that what tomorrow will happen. And the main thing is, one man also can analyze by a running a query on DBMS, whatever there will be some difference somehow good or bad. But if there you no having any information about your company, then what will happen to its future!!!

Disadvantages of DT and DM

Besides the huge advantages, there also a little bit of disadvantages. Like-

- *Security Issues.*

- *Privacy Issues.*

- *Incorrect Information.*

These noted issues are big issues, like; here all the issues are related to third party behavior. But, yes, they all can be maintained. If you are saving passwords and some physical securities or some rules that only a specified person only can check out that matters, then it will be very good to use these highly decorated and developed IT helps of DM and DT.

<div align="center">Conclusion and implications</div>

Whether business is big or small, it has to use database technology for the purpose of retrieving/maintaining relevant information. During recent times, this technology has been evolved significantly. Recent development of technology has unleashed the power of data mining and database. Construction of data ware house has gained importance for businesses. Big businesses have very high volume of data. Recent developments of software and hardware have improved the technique of data mining significantly. Data integration, data

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

migration and data summarization has become easy nowadays. Availability of parallel processing systems and high performing servers has changed the whole scenario.

Latest technological developments in database designing/developments and data mining have improved capabilities of businesses to understand data significantly. Quality of decision making has also been improved. It is easy to conduct data driven extensive research. At present many businesses are improving their data mining techniques. Little time is required to answer advance queries. Parallel processing is being used to explore large volume of data. Right approach in data mining helps businesses to keep them in focus. It is important for growth of business. Advance data modeling has become possible now. Prediction of data Trend/behavior data has become easy now. It is possible to perform statistical analysis in case of text data as well as web data.

NLP (Natural Language Processing) is used for easy analysis of data. Text may be turned into data. Analysis of text data is also carried out by opinion mining. It is also known as Sentiment analysis. It helps to do language processing. It further leads to information subtraction as well as getting information that is subjective. There are number of applications of Sentiment analysis i.e. social media and digital marketing. Here, consumer behavior can be easily judged. ADAPA and 11 Ants are examples of tools concerning data mining. Database coupling, Hadoop, big data etc. are additional tools. Machine learning approach and Statistical approaches are two different approaches as for as data mining techniques are concerned. Pre-existing data as well as secondary data are used in these approaches.

There are areas in which techniques of data mining are applicable i.e. e-commerce, finance, banking, biology, marketing etc. Now it is possible to do high quality data analysis,

DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

data interpretation, data prediction etc. It is possible to do data analysis that is multidimensional. Trends obtained from data may be analyzed for various purposes. Data analysis is used to provide quality service to customer as well as product promotion. Availability of easy internet and mobile technology i.e. mobile phone, email, web applications etc. are the reason behind growing popularity of data mining. There is significant use of data mining in the field of bioinformatics.

Data quality is the major concern in the field of data mining. Heterogeneity of data, matters of privacy, data uncertainty etc. are also major issues. Data interpretation, data modeling depends on reliability and precision of data. Data redundancy and data ambiguity are also major concerns. There should not be any uncertainty regarding data. End user should not be affected by complexities of data mining. Field of data mining is having bright future all over the world.
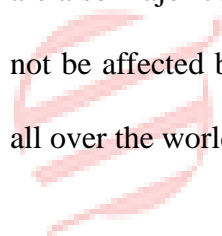
DATA MINING AND DATABASE TECHNOLOGY (WEB MINING, TEXT MINING, SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA, TOOLS, TECHNIQUES, METHODS, APPLICATIONS ETC.)

References

Fayadd, U., Piatesky -Shapiro, G., and Smyth, P. (1996). From Data Mining To Knowledge Discovery in Databases", The MIT Press, ISBN 0–26256097–6, Fayap, 1996.

Sharma, M. (2014) Data Mining: A Literature Survey, International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-3, Issue-2), http://www.ermt.net/docs/papers/Volume_3/2_February2014/V3N2-121.pdf

Huang, N.F, Ka, C.N, Hun, H.W., Jai, G.Y. and Lin, C.L. (2005). "Apply Data Mining to Defense-in-Depth Network Security System". Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA'05), 2005.

Gheware, S.D., Kejkar, A.S., Tondare, S.M. (2014), ISSN (Online): 2278-1021 International Journal of ... (n.d.).Retrieved from http://www.ijarcce.com/upload/2014/october/IJARCCE1C%20a%20santosh%20Data%20Mini ni

Related Topics :

Database Assignment Help

Data Analysis Assignment Help

Oracle Assignment Help