International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 3.45 (SJIF-2015), e-ISSN: 2455-2584
Volume 3, Issue 4, April-2017

Research Trends and its Applications in Data Mining

RachanaParikh[1],Rachit Adhvaryu[2]

[1]*Information Technology Department, Gujarat Technological University,*
[2]*Information Technology Department, Gujarat Technological University.*

Abstract—Knowledge or Information has played a significant role and had an immense impact on human activities since his development. Data mining is the process of knowledge discovery where knowledge is retrieved by learning the data stored in very large repositories, which are then analysed using various views and the result is précised into useful data or information. The feature of extracting knowledge/information from the large data repositories has made data mining a very important and definitearea of research affecting human life in directly or indirectly. The purpose of this paper is to provide the research trends in the field of data mining.

Keywords— Database, Data Mining, Information, Knowledge, Research

INTRODUCTION

Originated from Knowledge Discovery (KDD) from databases, also known as data Mining (DM), Data Mining (DM) is the task of discovering interesting patterns from large amounts of data where data can be stored in database, data ware house or other information repositories. Many techniques are currently used in this rapidly growing world, including statistical analysis and machine learning based approaches [1]. With the speedydevelopment of the internet and the large increase of unstructured databases, new technologies and applications are continuously being developed in this area.

The main challenges of data mining are:

• Data mining to deal with large amounts of data located at different sites. The amount of data that can easily exceed any limit.

• Data mining is very computationally intensive process involving very large databases. It is necessary to partition and distribute the data to reduce execution time and space and improve performance.

• Input data changesvery quickly. In many application domains, data to be mined either is produced with high rate or live streams. In those cases, knowledge has to be mined fast and efficiently in order to be usable and updated.[1]

DATA MINING OPPORTUNITIES

Data mining strives in searching for valuable business information in a large database — for example, finding relevant products from stored data — and mining a mountain to find valuable ones. Both processes require either surfing through an immense amount of material, or smartlyfinding exactly where the value resides [2]. Given the sets of databases of sufficient size and quality, data mining technology can generate new business opportunities by providing the following capabilities:

Automated prediction of trends and behaviours. Data mining automates the process of searching predictive data in large databases. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include weather forecasting. [2]

Automated discovery of previously unknown patterns. Data mining tools sweep through databases and identify previously hidden configurations in one step. An example of pattern discovery is the analysis of the either relevant or irrelevant products that are often purchased together. The most commonly used data mining techniques are [2]:

Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

Nearest neighbour method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset.

Rule induction: The extraction of useful if-then rules from data based on statistical significance.

DATA MINING FAMILY

1. Statistics

The most important member of data mining family is statistics. Without statistics, there would be no data mining, as statistics are the base of most of the technologies on which data mining is built. The concepts such as regression analysis,

standard distribution, standard deviation, standard variance, discriminate analysis, cluster analysis, and confidence intervals are used to study data and data relationships. These are the building blocks with which more statistical analysis isstrengthened. Today, classical statistical analysis plays a significant role [2].

2. Artificial Intelligence & Machine Learning

Data mining's second largest family member is artificial intelligence and machine learning. AI is built upon heuristics as opposed to statistics, and attempts to apply human-thought like processing to statistical problems. AI found few applications at very high end scientific/government markets which fulfilled the requirement of supercomputers [2]. Machine Learning could be considered as an evolution of AI, because it implements AI heuristics with advanced statistical methods. It lets the computer programs learn about the data they study and then apply learned knowledge to information [2].

3.Databases

Third family member is Databases. Huge amount of data needs to be stored in a repository or data sets, and that too needs to be managed in a proper way. Earlier data was managed in records and fields, then in various models like hierarchical, network etc. [2] Relational model proved to be more efficient data storage tool. But in data mining, volume of data is too high, so we need specialized servers for it. We call the term as Data Warehousing. Data warehousing also supports OLAP operations to be applied on it, to support decision making [2].

RESEARCH TRENDS AND ITS APPLICATIONS

The field of data mining has been growing rapidly due to its broad applicability, achievements and scientific progress, understanding. A number of data mining applications have been successfully implemented in various domains like fraud detection, retail, health care, finance, telecommunication, and risk analysis...etc.[3] Advancements in data mining with various integrations and implications of methods and techniques have shaped the present data mining applications to handle the various challenges.The areas in which data mining is applied are:

A. Fight against Terrorism: After 9-11 attacks, many countries imposed new laws against fighting terrorism. These laws allow intelligence agencies to effectively fight against terrorist organizations. USA launched Total Information Awareness program with the goal of creating a huge database of that consolidate all the information on population [3].

B. Bio-informatics and Cure for Diseases: The second most important application trend, deals with mining and interpretation of biological sequences and structures. Data mining tools are rapidly being used in finding genes regarding cure of diseases like Cancer and AIDS [3].

C. Web and Semantic Web: Web is the hottest trend now, but it is fully structured. Data mining is helping web to be organized, which is called Semantic web. The underlying technology is Resource Description Framework (RDF) which is used to describe resources. FOAF is also a supporting technology, heavily used in Facebook for tagging [3].

D. Business Trends: Today's business environment is more dynamic, so businesses must be able to react quicker, must be more profitable, and offer high quality services. Here, data mining serves as a fundamental technology in enabling customer's transactions more accurately, faster and meaningfully. Data mining techniques of classification, regression, and cluster analysis are used in business trends which make it intelligent [3].

Data mining has vast application scope and there are many fields where research can be made in great extent. The application areas are:

1.Healthcare

The past decade has seen an unbelievable growth in biomedical research, ranging from the development of new medicines to the identification and study of human genes. Recent research focuses on disease diagnosis, prevention and treatment [2].

2.Finance

Financial data collected in thebanking and financial industry is often relatively complete, reliable and high quality, which facilitates systematic data analysis and data mining [3].

3. Telecommunication

The telecommunication industry has quickly evolved from providing facilities to local and international levels. The integration of telecommunication, computer network, Internet and numerous other means of communication and computing is one of the significant achievements. This creates a great demand from data mining in order to help understand business involved, identify telecommunication patterns, catch fraudulent activities, make better use of resources ,and improve the quality of service [3].

4. Text Mining and Web Mining

Text mining is the process of finding large volumes of documents from certain keywords or key phrases. By searching literally thousands of documents various relationships between the documents can be established. Text mining can be used to derive certain patterns in the comments that may help to identify common perception via some types of surveys [3]. An extension of text mining is web mining. Web mining is a new field that integrates data and text mining within a website. It enhances the web site with intelligent behaviour, such as suggesting related links. They can be configured to check and collectinformation from a wide variety of locations and can analyse information across one or multiple sites. For example, the search engines work on the principle of data mining [3].

5. Distributed/Collective Data Mining

Much of the data mining is done on a database or data warehouse information which is resides in one place. However, the situation arises where information resides in different places, in different physical locations. This is known generally as distributed data mining (DDM) [3]. Therefore, the goal is to effectively mine distributed data which is located in heterogeneous sites. Distributed data mining (DDM) offers a different approach to traditional approaches analysis, by using a combination of localized data analysis, together with a global data model[3].

6. Ubiquitous Data Mining (UDM)

The invention of laptops, palmtops, cell phones, and wearable computers is making ubiquitous access to large quantity of data possible. Advanced analysis of data for extracting useful knowledge is the next natural step in the world of ubiquitous computing. Accessing and analysing data from a ubiquitous computing device offer many challenges [7]. The objective of UDM is to mine data while minimizing the cost of ubiquitous presence.Human-computer interaction is another challenging aspect of UDM. Visualizing patterns like classifiers, clusters, associations and others, in portable devices are usually difficult. The small display areas offer serious challenges to interactive data mining environments [7].

7. Hypertext and Hypermedia Data Mining

Hypertext and hypermedia data mining can be defined as mining data which includes text, hyperlinks, text mark-ups, and various other forms of hypermedia information. It is closely related to both web mining, and multimedia mining. Data mining techniques used for hypertext and hypermedia data mining include classification (supervised learning), clustering (unsupervised learning), semi-structured learning, and social network analysis [3].In the case of classification, or supervised learning, the process starts off by reviewing training data in which items are marked as being part of a certain class or group. This data is the basis from which the algorithm is trained. Unsupervised learning or clustering is concerned with the creation of hierarchies of documents based on similarity, and organize the documents based on that hierarchy. Semi-supervised learning is the case where there are both labelled and unlabelled documents, and there is a need to learn from both types of documents. Social network analysis is also applicable because the web is considered a social network, which examines networks formed through collaborative association. Graph distances and various aspects of connectivity come into play when working in the area of social networks [3].

8. Multimedia Data Mining

Multimedia Data Mining is the mining and analysis of various types of data, including images, video, audio, and animation. The main objective of multimedia data mining is mining data which contains different kinds of information.approach is to create a multimedia data cube which can be used to convert multimedia type data into a form which is suited to analysis using one of the main data mining techniques, but taking into account the unique characteristics of the data [5]. This may include the use of measures and dimensions for texture, shape, colour, and related attributes. Another developing area in multimedia data mining is audio data mining (mining music). The idea is mainly to use audio signals to indicate the patterns of data or to represent the features of data mining results [5].

9. Spatial and Geographic Data Mining

Spatial and Geographic data contains information about astronomical data, natural resources, or even orbiting satellites and spacecraft which transmit images of earth from out in space. Much of this data is image-oriented, and can represent a great deal of information if properly analysed and mined [4]. Analysing spatial and geographic data include such as understanding and browsing spatial data, uncovering relationships between spatial data items (and also between non-spatial and spatial items), and also analysis using spatial databases and spatial knowledge bases. The applications of these would be useful in such fields as remote sensing, medical imaging, navigation etc. Spatial and geographic data mining requires a lot of learning and analysis [4].

10. Time Series/Sequence Data Mining

It involves the mining of a sequence of data, which can either be referenced by time (time-series, such as stock market), or is simply a sequence of data which is ordered in a sequence. One approach of time series mining focuses on identifying movements or components such as seasonal variations [6].Another approach is Similarity Search; concerned with the identification of a pattern sequence which is close or similar to a given pattern, and this form of analysis can be broken down into two subtypes: whole sequence matching and subsequence matching. Whole sequence matching attempts to find all sequences which bear a likeness to each other, while subsequence matching attempts to find those patterns which are similar to a specified, given sequence. Sequential pattern mining has as its focus the identification of sequences which occur frequently in a time series or sequence of data. This is particularly useful in the analysis of customers, where certain buying patterns could be identified [6].

11. Constraint- Based Data Mining.

This form of data mining incorporates the use of constraints which guides the process. Frequently, this is combined with the benefits of multidimensional mining to add greater power to the process. There are several categories of constraints which can be used, each of which has its own characteristics and purpose [9]. These are:

Knowledge-type constraints. This type of constraint specifies the ―type of knowledge which is to be mined, and is typically specified at the beginning of any data mining query. Some of the types of constraints which can be used include clustering, association, and classification.

Data constraints. This constraint identifies the data which is to be used in the specific data mining query. Data constraints can be specified in a form similar to that of a SQL query.

Dimension/level constraints. Because much of the information being mined is in the form of a database or multidimensional data warehouse, it is possible to specify constraints which specify the levels or dimensions to be included in the current query.

Interestingness constraints. It would also be useful to determine what ranges of a particular variable or measures are considered to be particularly interesting and should be included in the query.

Rule constraints. It is also important to specify the specific rules which should be applied and used for a particular data mining query or application [9].

12. Phenomenal Data Mining

It focuses on the relationships between data and the phenomena which are inferred from the data [8]. One example of this is that using receipts from cash supermarket purchases, it is possible to identify various aspects of the customers who are making these purchases. Some of these phenomena could include age, income, ethnicity, and purchasing habits. One aspect of phenomenal data mining is the need to have access to some facts about the relations between these data and their related phenomena [8].

CONCLUSIONS

In this paper wehave presented the various research areas in data mining and its applications. Above literature focuses on promising areas of data mining. Though very limited areas are described here in this paper, yet there are a lot more areas in which research can be carried out to great extent. This paper provides a new approachforthe researchers to work on various data mining areas and develop systems or tools useful for the society.

REFERENCES

Heikki, Mannila, "*Data mining: machine learning, statistics, and databases*", Statistics and Scientific Data Management, pp. 2-9. 1996.

Han, J. and M. Kamber, "*Data Mining: Concepts and Techniques*", Morgan Kaufmann, 2001.

Annan Naidu Paidi, "*Data Mining: Future Trends and Applications*", International Journal of Modern Engineering Research (IJMER), Vol.2, Issue.6, Nov-Dec. 2012 pp-4657-4663

Miller and J. Han (eds.), "*Geographic Data Mining and Knowledge Discovery*", Taylor and Francis, 2001.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., "*Multimedia mining*", SEAS Transactions on Systems, No 3, s. 3263-3268, 2005.

Huysmans, Baesens, Martens, Denys and Vanthienen, "*New Trends in Data Mining"*, TijdschriftvoorEconomie en Management, vol. L, 4, 2005.

Salmin, Sultana et al., "*Ubiquitous Secretary: A Ubiquitous Computing Application Based on Web Services Architecture*", InternationalJournal of Multimedia and Ubiquitous Engineering Vol. 4, No. 4, October, 2009.

Jing He, "*Advances in Data Mining: History and Future"*, Third international Symposium on Information Technology Application, 978-0-7695-3859-4/09 IEEE 2009 DOI 10.1109/IITA.2009.204.

Venkatadari M., Dr. Lokanataha C. Reddy, "*A Review on Data Mining From Past to Future"*, International Journal of Computer Applications, pp.19-22, vol. 15, No. 7, Feb 2011.