Data Mining Applications and Feature Scope Survey

Dr Shahid[1], Mr Bijay Singh[2], Shuchi Sethi[3]

[1]AP ,Deptt of ICT, ISBAT University, Kampala Uganda

[2]Astt Prof, Netaji Subhash Institute of Technology, Patna

[3]Research Scholar, Dept of Computer Science, Jamia Millia Islamia, New Delhi

Abstract: *We have concentrated on a range of strategies, methodologies, and distinct fields of research in this article, all of which are useful and relevant in the field of data mining technologies. As we all know, numerous multinational corporations and major corporations operate in various parts of the world. Each location of business may create significant amounts of data. Corporate decision-makers need access to all of these data sources in order to make strategic decisions. The data warehouse adds substantial value to the firm by increasing the efficiency of management decision-making. The significance of strategic information systems like these is immediately recognised in an uncertain and highly competitive corporate climate, but in today's business world, efficiency or speed is not the sole route to competitiveness. This massive amount of data is available in the form of terabytes to petabytes, which has profoundly impacted research and engineering. To evaluate, manage, and make decisions with such a large volume of data, we need data mining tools, which will alter numerous fields. This work provides a greater number of data mining applications as well as a more focused scope of data mining, which will be useful in future research.*

Keywords: Task data mining and web mining, Life cycle in data mining, data mining visualization, Application on data mining.

1. Introduction

Because the data is available in a variety of formats, the appropriate action may be made. Not only should these facts be analysed, but they should also be used to make excellent decisions and keep track of them. The data should be obtained from the database as and when the client requires it in order to make the best decision possible. This method is referred to as data mining, knowledge hub, or simply KDD (Knowledge Discovery Process). The finding of helpful the perception of "we are data abundant but information poor" drew a lot of attention in the field of information technology.

Due to knowledge from massive collections of data in the subject of "Data mining,"

There is a massive amount of data, but we are hardly able to transform it into meaningful information and knowledge for corporate decision-making. It is necessary to collect a large amount of data in order to develop information. Different media, such as audio/video, numbers, text, figures, and hypertext formats, may be used. To fully use data, a tool for automatic data summarization, extraction of the core of stored information, and pattern detection in raw data is required.

With the massive amounts of data saved in files, databases, and other repositories, it is becoming increasingly vital to build effective tools for data analysis and interpretation, as well as the extraction of useful information that may aid decision-making.

The one and only Data Mining' is the answer to all of the above. The extraction of hidden predictive data is known as data mining. Information from enormous datasets; it's a strong tool with a lot of promise for helping people. In

their data warehouses, firms concentrate on the most important information [1,2,3,4]. Data mining software forecasts future patterns and behaviours, allowing businesses to take preventative measures. Decisions based on knowledge [2]. Data mining's automated, prospective assessments are a game changer.

Beyond the analysis of previous occurrences offered by prospective decision-making tools,
systems. Data mining techniques can provide answers to queries that were previously too time consuming to address. it takes a long time to fix They create databases in order to uncover hidden patterns and make predictions. Information that specialists may overlook because it falls outside of their usual scope.

We presented a new approach of defining the KDD Process. Section 6 provides a brief overview of some of the most often used data mining techniques. The heart of the article is Chapter 7, in which we examine applications and recommend feature directions for various data mining applications.

1.1 Data Analysis for Exploratory Purposes:
A tremendous quantity of information is available in the repositories. This data mining activity will accomplish two goals (i). Without knowing what the consumer is looking for, it (ii) analyses the data.
For the client, these tactics are engaging and visible.

1.2 Modeling that is descriptive:
It contains models for the data's overall probability distribution, partitioning of the p-dimensional space into groups, and models characterising the connections between the variables.

1.3 Modeling for Prediction:
This approach allows the value of one variable to be predicted based on the values of other variables that are known.

1.4 Patterns and Rules to Look for:
This assignment is mostly utilised to uncover the cluster's hidden pattern as well as to locate the hidden pattern. A cluster has a variety of designs and clusters of various sizes. The goal of this work is to figure out "how best we can recognise patterns." This may be performed by employing rule induction and other data mining approaches such as (K-Means/K-Medoids). This is referred to as the clustering algorithm.

1.5 Content-based retrieval:
The main goal of this work is to locate data sets that are regularly utilised in the audio/video and picture fields. It is the discovery of a pattern in the data set that is comparable to the pattern of interest.

2. Data Mining System Types:
A variety of characteristics may be used to classify data mining systems. The categorisation is as follows:

2.1 Life Cycle of Data Mining:
A data mining project's life cycle is divided into six stages[2,4]. The stages are not in any particular order. It's constantly necessary to switch back and forth
between stages. It is determined by the results of each step. The following are the key stages:

2.1.1 Understanding of Business:
This phase focuses on collecting a business knowledge of the project objectives and requirements, then translating that information into a data mining issue definition and a preliminary plan to achieve the goals.
2.1.2 Data comprehension:
It begins with a data gathering phase to familiarise yourself with the data, find data quality issues, get early insights into the data, or identify intriguing subsets to generate hypotheses about hidden information.
2.1.3 Preparation of Data:
This step takes all of the different data sets and creates the different types of activities based on the raw data.

2.2 Data Mining Model Visualization

The basic goal of data visualisation is to convey the general concept of the data mining methodology. The majority of the time in data mining, we are getting data from repositories that are concealed. For a user, this is the most challenging task. As a result, this depiction of the data mining approach aids us in providing the highest levels of comprehension and trust.

Clustering is a phrase that refers to analysing various data items without consulting a recognised class level. Unsupervised learning or segmentation are other terms for it. It is the process of dividing or segmenting data into groups or clusters. Domain specialists evaluate the behaviour of the data to determine the clusters. The phrase segmentation has a very precise meaning; it refers to the division of a database into separate groups of comparable tuples. The process of displaying the summarised information from the data is known as summarization. The association rule determines the relationship between the various properties. The mining of association rules is a two-step procedure:

Identifying all frequent item sets and generating strong association rules from them.

Table 1.1 to describe the new form the word

Data	A set of facts, F.
Pattern	An expression E in a language L describing facts in a subset F_E of F.
Process	It means different operations associated with the KDD .The operations involving preparation of the data ,searching the different patterns , Judging the knowledge and evaluation etc.
Valid	Those patterns which are discovered that are completely new one and which can be used feature
Novel	Derive the hidden patterns
Useful	Newly discovered patterns should be used for different actions .

3. Methods of Data Mining:
Rules and Decision Trees
Methods of Nonlinear Regression and Classification
Methods based on Examples
Graphical Dependency Models with Probabilistic Constraints
Models of Relational Learning

4. Applications of Data Mining in Healthcare:
Health data mining applications have a lot of promise and can be very beneficial .However, the availability of good healthcare data is critical to the success of healthcare data mining. In this regard, the healthcare business must investigate how data may be acquired, saved, processed, and mined more effectively. Standardization of clinical language and data sharing across companies are two possible routes for enhancing the advantages of healthcare data mining technologies.

4.1 Data mining is used for market basket analysis:
MBA students employ data mining techniques (Market Basket Analysis). When a consumer wants to buy anything, this approach aids us in determining the relationships between the many goods that the customer has placed

in their shopping carts. The finding of such relationships, which enhances the business technique, may be found here. In this approach, merchants employ data mining techniques to determine which consumers' intentions are (buying the different pattern). In this way, the strategy is employed to increase business revenues while also assisting in the purchase of connected things.

5. Data Mining's Purpose

Searching for important business information in a vast database, for example, locating connected goods in terabytes of store scanner data, and mining a mountain for a vein of lucrative metal are all examples of data mining. Both techniques need either sorting through a massive amount of data or probing it intelligently to determine where the value is hidden. Data mining technology, when used with datasets of appropriate size and quality, can open up new business prospects by enabling the following capabilities:

6. Conclusion:

The numerous data mining applications were briefly explored in this study. This review will aid academics in concentrating on the many aspects of data mining. In a future course, we'll look at several classification techniques and the importance of using evolutionary computing (genetic programming) to create effective data mining classification systems. The majority of earlier research on data mining applications in various industries used a wide range of data kinds, from text to pictures, and stored them in a variety of databases and data structures. Different data mining approaches are employed to extract patterns and hence knowledge from these various datasets. Data and technique selection for data mining is a crucial responsibility in this process, and it necessitates understanding. Several attempts have been made to design and build a generic data mining system, but no system has been found to be completely universal. As a result, for each domain, a domain expert's assistant is necessary. Domain experts will lead the domain experts to successfully use their experience toward the production of data mining system knowledge Domain specialists must determine the type of data that should be collected in a specific issue area, as well as the selection of specific data for data mining, as well as the cleansing of data and data processing, pattern extraction for knowledge development, and pattern interpretation and knowledge development.

Reference

[1] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

[2] Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, ohn Wiley & Sons, Inc, 2005.

[3] Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education,New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006

[4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R... "CRISP-DM 1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen Bank Group B.V (The Netherlands), 2000".

[5] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge

[6] Discovery in Databases," AI Magazine, American Association for Artificial Intelligence, 1996.

[7] Tan Pang-Ning, Steinbach, M., Vipin Kumar. "Introduction to Data Mining", Pearson Education, New Delhi, ISBN: 978-81-317-1472-0, 3rd Edition, 2009. Bernstein, A. and Provost, F., "An Intelligent Assistant for the Knowledge Discovery Process", Working Paper of the Center for Digital Economy Research, New York University and also presented at the IJCAI 2001 Workshop on Wrappers for Performance Enhancement in Knowledge Discovery in Databases.

[8] Baazaoui, Z., H., Faiz, S., and Ben Ghezala, H., "A Framework for Data Mining Based Multi-Agent: An Application to Spatial Data, volume 5, ISSN 1307-6884," Proceedings of World Academy of Science, Engineering and Technology, April 2005.

[9] Rantzau, R. and Schwarz, H., "A Multi-Tier Architecture for High-Performance Data Mining, A Technical Project Report of ESPRIT project, The consortium of CRITIKAL project, Attar Software Ltd. (UK), Gehe AG (Denmark); Lloyds TSB Group (UK), Parallel Applications Centre, University of Southampton (UK), BWI, University of Stuttgart (Denmark), IPVR, University of Stuttgart (Denmark)".

[10] Botia, J. A., Garijo, M. y Velasco, J. R., Skarmeta, A. F., "A Generic Data mining System basic design and implementation guidelines", A Technical Project Report of

[11] CYCYTprojectofSpanishGovernment.1998.WebSite: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.1935

[12] Campos, M. M., Stengard, P. J., Boriana, L. M., "Data-Centric Automated Data Mining",WebSite.:www.oracle.com/technology/products/bi/odm/pdf/automated_data_mining_paper_1205.pdf.

[13] Amit ,Choudhary S P Singh, V K Pandey; 'A Low Power and High Gain CMOS Tunable OTA with Cascade Current Mirrors',Volume No.2,Issue No.1,2013,PP.075-078,ISSN :2229-5828

[14] Anju Gauniya Pandey , Sanjita Das , S. P.Basu, Palak Srivastava; 'Design and Evaluation Of Nanoemulsion For Delivery of Diclofenac Sodium',Volume No.2,Issue No.1,2013,PP.079-082,ISSN :2229-5828

[15] Raj Kumar Goel , Rinku Sharma Dixit, Dr. Manu Pratap Singh; 'Implementaion of Pattern Storage Neural network As Associative Memory For Storage and Recalling of Finger Prints',Volume No.2,Issue No.1,2013,PP.083-090,ISSN :2229-5828

[16] Amit Kumar Yadav, Satyendra Sharma; 'Design and Simulation of Multiplier for High -speed Application',Volume No.2,Issue No.2,2014,PP.001-007,ISSN :2229-5828

[17] Deepak Kumar ,Anjana Rani Gupta, Somesh Kumar; 'Dynamic Simulation of Multiple Effect Evaporators in Paper Industry Using MATLAB',Volume No.2,Issue No.2,2014,PP.008-014,ISSN :2229-5828

[18] Devendra Pratap, Satyendra Sharma; 'Planning and Modelling of Indoor WLAN Through Field Measurement at 2.437 GHz Frequency',Volume No.2,Issue No.2,2014,PP.015-019,ISSN :2229-5828