The Criticism of Data Mining Applications and Methodologies

Amiya Kumar Sahu
Research Scholar, JJT University,
Jhunjhunu(Rajasthan) INDIA

Abstract: In this paper we have attentive a diversity of procedures, methodologies and dissimilar parts of the exploration which are cooperative and noticeable as the significant crushed of data mining Technologies. The data warehouse is used in the significant business value by improving the effectiveness of managerial decision-making. In an uncertain and highly competitive corporate environment, the worth of planned information systems such as these are easily documented though in today's commercial environment, competence or rapidity is not the merely key for affordability.

1. INTRODUCTION

Each and every day the human beings are using the vast data and these data are in the different fields .It may be in the form of documents, may be graphical formats ,may be the video ,may be records (varying array).As the data are available in the different formats so that the proper action to be taken. Not only to analyze these data but also take a good decision and maintain the data. As and when the customer will require the data should be retrieved from the database and make the better decision .This technique is actually we called as a data mining or Knowledge Hub or simply KDD (Knowledge Discovery Process). There is huge volume of data but we hardly able to turn them in to useful information and knowledge for managerial decision making in business. It's going to show a discrepancy formats like audio/video, numbers, text, figures, machine-readable text formats. to require complete advantage of data; the information retrieval is just not enough, it needs a tool for automatic account of information, extraction of the essence of knowledge keep, and also the discovery of patterns in data. With the big quantity of information keep in files, databases, and different repositories, it's progressively necessary, to develop powerful tool for analysis and interpretation of such knowledge and for the extraction of attention-grabbing information that might facilitate in decision-making. The sole answer to any or all on top of is 'Data Mining'. Data processing is that the extraction of hidden prognostic info from massive knowledgebase's; it's a strong technology with nice potential to assist organizations target the foremost necessary info in their data warehouses. Data processing tools predict future trends and behaviors, helps organizations to form proactive knowledge-driven choices [2]. The automatic, prospective analyses offered by data processing move on the far side the analyses of past events provided by prospective tools typical of call support systems. Data processing tools will answer the queries that historically were too time intense to resolve.

2. Data mining Tasks

The two "high-level" major goals of data mining, in practice, are *prediction* and *description*. **Prediction** involves by means of some variables or fields in the database to forecast unknown or future values of other variables of interest. **Description** focuses on discovery human-interpretable patterns relating the data.The goals of prediction and description are achieved by using the following most important **data mining tasks**:

2.1 Exploratory Data Analysis: In the repositories vast amount of information's are available. This data mining task will serve the two purposes: (i).Without the knowledge for what the customer is searching, and then (ii) It analyzes the data and techniques which are interactive and visual to the customer.

2.2 Prophetic Modeling: This model permits the worth of 1 variable to be foretold from the famous values of different variables.

2.3. Discovering Patterns and Rules: This task is primarily wont to realize the hidden pattern also on discover the pattern within the cluster. During a cluster variety of patterns of various size and clusters area unit accessible .The aim of this task is "how best we'll find the patterns".

2.4 Retrieval by Content: The first objective of this task is to seek out knowledge the info the information sets of times employed in the for audio/video also as pictures it's finding pattern almost like the pattern of interest within the data set.

3. Forms of data processing System: Data processing systems are categorized per numerous criteria the classification is as follows [3]:

3.1 Classification information of knowledge: In a corporation a large quantity of data's are accessible wherever we want to classify these data however these are accessible most of times during a similar fashion. We want to classify this information per its type (maybe audio/video, text format etc)

3.2 Classification information of knowledge: There are such a lot of variety of knowledge mining models (relative data model, Object Model, Object orientating information Model, stratified information Model/W information model) are accessible and every and each model we have a tendency to are exploitation the various information. According to these information model the information mining system classify the information within the model.

3.3 Classification of information mining systems in keeping with the type of data discovered: This classification supported the type of data discovered or data processing functionalities, like characterization, discrimination, association, classification, clustering, etc. Some systems tend to be comprehensive systems giving many data processing functionalities along.

3.4 Classification of information mining systems in keeping with mining techniques used: This classification is in keeping with the information analysis approach used like machine learning, neural networks, genetic algorithms, statistics, visualization, info homeward or information warehouse-oriented, etc. The classification may also take under consideration the degree of user interaction concerned within process like query-driven systems, interactive beta systems, or autonomous systems.

4. Data Mining Applications

In this section, we've got centered a number of the applications of information mining and its techniques are analyzed severally Order.

4.1 Data processing Applications in health care data processing applications in health will have tremendous potential and quality: However, the success of health care data processing hinges on the supply of unpolluted health care knowledge. During this respect, it's vital that the health care business look at however knowledge will be higher captured, stored, ready and mined. Doable directions embrace the standardization of clinical vocabulary and therefore the sharing of information across organizations to reinforce the advantages of health care data processing applications.

4.2 Future Directions of Health care system through data processing Tools: As health care knowledge don't seem to be restricted to merely quantitative knowledge (e.g., doctor's notes or clinical records), it's necessary to additionally explore the utilization of text mining to expand the scope and nature of what health care data processing will presently do. This is often especially wont to mix all the information then mining the text. It's additionally helpful to seem into however pictures (e.g., imaging scans) will be brought into health care data processing applications. It's noted that progress has been created in these areas.

4.3 Data mining is used for market basket analysis Data mining technique is used in MBA(Market Basket Analysis): When the customer want to buying some products then this technique helps us finding the associations between different items that the customer put in their shopping buckets. Here the discovery of such associations that promotes the business technique .In this way the retailers uses the data mining technique so that they can identify that which customers intension (buying the different pattern).In this way this technique is used for profits of the business and also helps to purchase the related items.

4.4 The information mining is employed a rising trends within the education system: Because the numbers of universities are established aspect by aspect, each day a millennium of scholar's are enrolls across the country. With immense range of upper education aspirants, we have a tendency to believe that data processing technology will facilitate bridging information gap in higher academic systems. The hidden patterns, associations, and anomalies that are discovered by data processing techniques from academic knowledge will improve deciding processes in higher academic systems. This improvement will bring blessings like increasing academic system potency, decreasing student's drop-out rate, and increasing student's promotion rate, increasing student's retention rate in, increasing student's transition rate, increasing academic improvement quantitative relation, increasing student's

success, increasing student's learning outcome, and reducing the price of system processes.

4.5 Data mining is now used in many different areas in manufacturing engineering: When we retrieve the data from manufacturing system then the customer is to use these data for different purposes like to find the errors in the data, to enhance the design methodology, to make the good quality of the data, how best the data can be supported for making the decision. But most of time the data can be first analyzed then after find the hidden patterns which will be control the manufacturing process which will further enhance the quality of the products .Since the importance of data mining in manufacturing has clearly increased over the last 20 years, it is now appropriate to critically review its history and Application.

4.6Future Directions in the manufacturing Engineering: Generally when we mine the data in the manufacturing, we do not give more important to the quality of the rules. After mining those knowledge which has generated is very difficult because relationship identification is too complex to understand. That's why we need the further to enhance the research methodology to know the proper knowledge. The new methodology was proposed i.e CRISP-DM which will provides the high level detail steps of instructions for using the data mining in the engineering field.

4.7 Data Mining Applications can be generic or domain specific: Some generic data mining applications cannot take its own these decisions but guide users for selection of data, selection of data mining method and for the interpretation of the results. The multi agent based data mining application [8, 10] has capability of automatic selection of data mining technique to be applied. The Multi Agent System used at different levels [8]: First, at the level of concept hierarchy definition then at the result level to present the best adapted decision to the user. This decision is stored in knowledge Base to use in a later decision-making. Multi Agent System Tool used for generic data mining system development [10] uses different agents to perform different tasks.

4.8 A multi-tier data mining system is proposed to enhance the performance of the data mining process [9]: It has basic components like user interface, data mining services, data access services and the data. There are three different architectures presented for the data mining system namely one-tire, Two-tire and Three-tier architecture. Generic system required to integrate as many learning algorithms as possible and decides the most appropriate algorithm to use. CORBA (Common Object Request Broker Architecture) has features like: Integration of different applications coded in any programming language considerably easy. It allows reusability in a feasible way and finally it makes possible to build large and scalable system. The data mining system architecture based on CORBA is given by Object Management Group [10] has all characteristics to accomplish a distributed and object oriented computation. A data-centric focus and automated methodologies makes data mining accessible to no experts [11].

4.9 Application of Data mining techniques in CRM: It aims to give a research summary on the application of data mining in the CRM domain and techniques which are most often used. Although this review cannot claim to be exhaustive, it does provide reasonable insights and shows the incidence of research on this subject. The results

presented in this paper have several important implications: Research on the application of data mining in CRM will increase significantly in the future based on past publication rates and the increasing interest in the area. The majority of the reviewed articles relate to customer retention.

4.10 The Domain Specific Applications: The domain specific applications are focused to use the domain specific data and data mining algorithm that targeted for specific objective. The applications studied in this context are aimed to generate the specific knowledge. In the different domains the data generating sources generate different type of data. Data can be from a simple text, numbers to more complex audio-video data. To mine the patterns and thus knowledge from this data, different types of data mining algorithms are used. The collection and selection of context specific data and applying the data mining algorithm to generate the context specific knowledge is thus a skillful job.

4.11 In language research and language engineering much time extra linguistic information is needed about a text: A linguistic profile that contains large number of linguistic features can be generated from text file automatically using data mining [14]. This technique found quite effective for authorship verification and recognition. A profiling system using combination of lexical and syntactic features shows 94% accuracy in selecting correct author for the text. The linguistic profiling of text effectively used to control the quality of language and for the automatic language verification.[14] This method verifies automatically the text is of native quality. The results show that language verification is indeed possible.

4.12 In Medical Science, there is large scope for application of data mining: Diagnosis of diesis, health care, patient profiling and history generation etc. are the few examples. Mammography is the method used in breast cancer detection. Radiologists face lot of difficulties in detection of tumors that's why CAM (Computer Aided Methods) could helps to the medical staff, so that they can produce the good quality of the result detection [14]. The neural networks with back-propagation and association rule mining used for tumor classification in mammograms. The data mining effectively used in the diagnosis of lung abnormality that may be cancerous or benign [14]. A REMINDS (Reliable Extraction and Meaningful Inference from Non-structured Data) system [21] integrates the structured and unstructured clinical data in patient records to automatically create high quality structured clinical data. To adopt the high quality technique, we can mine the existing patient records to support guidelines and give compliance to improve patient care. [21]

4.13 Data Mining methods are used in the Web Education: Data mining methods are used in the web Education which is used to improve courseware. The relationships are discovered among the usage data picked up during students' sessions. This knowledge is very useful for the teacher or the author of the course, who could decide what modifications will be the most appropriate to improve the effectiveness of the course. Web Education is rapidly growth in the application of data mining methods to educational chats which is both feasible and can be improvement in learning environments in the 21st century.

4.14 The Intrusion Detection in the Network: The intrusion detection in the Network is very difficult and needs a very close watch on the data traffic. The intrusion detection plays an essential role in computer security. The classification method of data mining is used to classify the network traffic normal traffic or abnormal traffic.[24]. If any TCP header does not belong to any of the existing TCP header clusters, then it can be considered as anomaly.

4.15 A malicious Executable is Threat: A malicious executable is threat to system's security, it damage a system or obtaining sensitive information without the user's permission. The data mining methods used to accurately detect malicious executables before they run [24]. Classification algorithms RIPPER, Naive Bays, and a Multi-Classifier system are used to detect new malicious executables. This classifier had shown detection rate 94.44%.

4.16 Sports Data Mining: Data mining is not only use in the business purposes but also it used in the sports .In the world, a huge number of games are available where each and every day the national and international games are to be scheduled, where a huge number of data's are to be maintained .The data mining tools are applied to give the information as and when we required, the open source data mining tools like WEKA and RAPID MINER frequently used for sport. This means that users can run their data through one of the built-in algorithms, see what results come out, and then run it through a different algorithm to see if anything different stands out.

4.17 The Intelligence Agencies collect and analyze information to investigate terrorist activities: Now a days the intelligence agency are using the sophisticated data mining algorithms which makes it easy, to handle the very large data bases databases for organizations..Though the organization's have used large data bases but data mining helps us to generate the different types of information in the organization like personal details of the persons along with, vehicle details .In data mining the Clustering techniques is used (Association rule mining) for the different objects(like persons, organizations, vehicles etc.) in crime records. Not only data mining detects but also analyzes the crime data. The classification technique is also used to detect email spamming and also find person who has given the mail. String comparator is used to detect deceptive information in criminal record.

4.18 The data mining system implemented at the Internal Revenue Service: The major lines of investigation included visualization of the relationships and data mining to identify and rank possibly abusive tax avoidance transactions. To enhance the quality of product data mining techniques can be used effectively. The data mining technology SAS/EM is used to discover the rules those are unknown before and it can improve the quality of products and decrease the cost. A regression model and the neural network model when applied for this purpose given accuracy above 80%. The neural network model found better than the regression model.

4.19 E-commerce is additionally the foremost prospective: The combination of e-commerce and data processing considerably improve the results and guide the users in generating information and creating correct business selections. This integration effectively solves many major issues related to horizontal data processing tools as well as the big effort needed in pre-processing of the info before it are often used for mining, and creating the results of mining unjust.

4.20 The Digital Library : Retrieves the info mining application are often employed in the sector of the Digital Library wherever the user can finds or collects, stores and preserves the info that square measure within the type of digital mode. The appearance of electronic resources and their magnified use in libraries has caused important changes in Library. The info and knowledge square measure on the market within the completely different formats. These formats embrace Text, Images, Video, Audio, Picture, Maps, etc. thus digital library could be an appropriate domain for application of knowledge mining.

4.21 The prediction in engineering applications was treated effectively by a data mining approach [14]: The prediction problems like the cost estimation problem in engineering, the problem of engineering design that involves decisions where parameters, actions, components, and so on are selected. Data mining technique is used for the variety of the parameters in the field of engineering applications like prior data .Once we gather the data then we can generate the different models, algorithms which will predict to different characteristic.

5. Conclusion

In this paper we momentarily studied the numerous data mining applications. This assessment would be cooperative to academics to emphasis on the numerous subjects of data mining. In upcoming sequence, we will appraisal the countless classification algorithms and significance of evolutionary computing (genetic programming) approach in manipulative of well-organized classification algorithms for data mining.

REFERENCES

[1] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892094-02-4, Two Crows Corporation, 10400 Falls Road, Potomac, MD 20844 (U.S.A.), 1999.

[2] Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining".

[3] Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-4448-484-4, 1st Edition, 2004

[4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R... "CRISP-DM1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA andDenmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen BankGroup B.V (The Netherlands), 2000".

[5] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge Discovery in Databases," AI Magazine, American Association for Artificial Intelligence, 1994.

[6] Tan Pang-Ning, Steinbach, M., Vipin Kumar. "Introduction to Data Mining", Pearson Education, New International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, June

[7] Baazaoui, Z., H., Faiz, S., and Ben Ghezala, H., "A Framework for Data Mining Based Multi-Agent:An Application to Spatial Data, volume 4, ISSN 1304-4884," Proceedings of World Academy of Science, Engineering and Technology, April 2004.

[8] Rantzau, R. and Schwarz, H., "A Multi-Tier Architecture for High-Performance Data Mining.

[9] 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 CYCYTprojectofSpanishGovernment.1998.

[10] 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_1204.pdf

[11] Sirgo, J., Lopez, A., Janez, R., Blanco, R., Abajo, N., Tarrio, M., Perez, R., "A Data Mining Engine based on Internet, Emerging Technologies and Factory Automation.

[12] Bianca V. D.,PhilippeBoula de Mareüil and Martine Adda-Decker, "Identification of foreign-accented French using data mining techniques, Computer Sciences Laboratory for Mechanics and Engineering Sciences (LIMSI)".Website

[13] Bianca V. D.,PhilippeBoula de Mareüil and Martine Adda-Decker, "Identification of foreign-accented French using data mining techniques, Computer Sciences Laboratory for Mechanics and Engineering Sciences (LIMSI)".

[14] Halteren, H. V., Oostdijk N., "Linguistic profiling of texts for the purpose of language verification,The ILK research group, Tilburg centre for Creative Computing and the Department of Communication and Information Sciences of the Faculty of Humanities,

[15] Antonie, M. L., Zaiane, O. R.,Coman, A., "Application of Data Mining Techniques for MedicalImage Classification", Proceedings of the Second International Workshop on Multimedia Data Mining MDM/KDD 2001) in conjunction with ACM SIGKDD conference, San Francisco, August 24,2001.

[16] Kusiak, A., Kernstine, K.H., Kern, J.A., McLaughlin, K.A., and Tseng, T.L., "Data Mining: Medical and Engineering Case Studies". Proceedings of the Industrial Engineering Research 2000

[17] Luis, R., Redol, J., Simoes, D., Horta, N., "Data Warehousing and Data Mining System Applied to E Learning, Proceedings of the II International Conference on Multimedia and Information& Communication Technologies in Education, Badajoz, Spain, December 3-4th 2003.

[18] Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., Zheng, R., Atabakhsh, H., Crime Data Mining: An Overview and Case Studies", A project under NSF Digital Government Programme,USA, "COPLINK Center: Information and Knowledge Management for Law Enforcement,", July2000 -June 2003

[19] Kay, J., Maisonneuve, N., Yacef, K., Zaiane O., "Mining patterns of events in students' teamwork data", Proceedings of the ITS (Intelligent Tutoring Systems) 2004 Workshop on Educational Data Mining, pages 44-42, Jhongli, Taiwan, 2004.

[20] Rao, R. B., Krishnan, S. and Niculescu, R. S., "Data Mining for Improved Cardiac Care" , SIGKDD Explorations Volume 8, Issue 1.

[21] Ghani, R., Probst, K., Liu, Y., Krema, M., Fano, A., "Text Mining for Product Attribute Extraction",SIGKDD Explorations Volume 8, Issue 1.

[22] DeBarr, D., Eyler-Walker, Z., "Closing the Gap: Automated Screening of Tax Returns to Identify Egregious Tax Shelters". SIGKDD Explorations Volume 8, Issue 1.

[23] Kanellopoulos, Y., Dimopulos, T., Tjortjis, C., Makris, C. "Mining Source Code Elements for Comprehending Object-Oriented Systems and Evaluating Their Maintainability", SIGKDD, Explorations Volume 8, Issue 1. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, June 2012.