Applications, Techniques and Trends of Data Mining and Knowledge Discovery Database

Khin Sein Hlaing[1], Yin Myo Kay Khine Thaw[2]

[1]Faculty of Information Science Department,
[2]Information Technology Supporting and Maintenance Department,
[1,2]University of Computer Studies, Mandalay, Myanmar

ABSTRACT

Data Mining and Knowledge Discovery is intended to be the best technical publication in the field providing a resource collecting relevant common methods and techniques. Traditionally, data mining and knowledge discovery was performed manually. As time passed, the amount of data in many systems grew to larger than terabyte size, and could no longer be maintained manually. Besides, for the successful existence of any business, discovering underlying patterns in data is considered essential. This paper proposed about applications, techniques and trends of Data Mining and Knowledge Discovery Database.

KEYWORDS: *Rule Data Mining, Knowledge Discovery in Data base.*

I. INTRODUCTION

Data Mining and Knowledge Discovery in Databases (KDD) is a rapidly developing area of research and application that builds on techniques and theories from many fields including statistics databases pattern recognition and learning data visualization uncertainty modelling data warehousing and Online Analytical Processing (OLAP) optimization and high performance computing. Data mining is used to find or generate new useful information's from large amount of data base. It is a process of extracting previously unknown and process able information from large databases and using it to make important business decisions. [2]

Data mining is a process of extracting previously unknown and process able information from large databases and using it to make important business decisions. It is also called as knowledge discovery process, Data mining should be used exclusively for the discovery stage of the KDD process. [3]. Data Mining Methods that including classification clustering probabilistic modelling prediction and estimation dependency analysis search and optimization. KDD is concerned with issues of scalability the multi-step knowledge discovery process for extracting useful patterns and models from raw data stores of including data cleaning and noise modelling and issues of making discovered patterns understandable. [2]

II. DATA MINING

Data mining is the process of discovering actionable information from large sets of data [4]. Data mining uses mathematical analysis to derive patterns and trends that exist in data. These patterns and trends can be collected and defined as a data mining model. [2].

Data Mining Trends: The field of data mining has been growing due to its great success in terms of broad-ranging application accomplishments and scientific progress, understanding. Advancements in data mining with several consolidations and implications of methods and techniques have molded the present data mining applications to handle the several challenges, the current trends of data mining applications are Research and Scientific Computing Trends .The explosion in the amount data from many scientific disciplines, such as astronomy, remote sensing, Bioinformatics, combinatorial chemistry, medical imagery, and experimental physics is moving to several data mining techniques, to find out useful information. The Direct-kernel based techniques a potential data mining tool for prognostic modeling, feature selection and visualization in scientific computing. Most of the current business data mining applications utilize the classification and prediction techniques for supporting business decisions. In business environment data mining has evolved to Decision Support Systems (DSS) and very lately it has grown to Business Intelligence (BI) systems. [5]

Due to the day to day change in technology the data mining trends are also affected by the change in technology because new techniques are very useful for the data and mining as well as for improving the old results. Referable to the enormous success of several application areas of data mining, the field of data mining has established itself as the major discipline of computer science and has shown interest potential for the future evolutions. Ever increasing technology and future application areas are always present new challenges and opportunities for data mining, the typical future trends of data mining includes Extraction and

preprocessing of data, Complex objects of data, Computing resources, Web mining, Scientific Computing and Business data. [5]

III. Data Mining Techniques
Data mining is highly effective, some of the data mining techniques are

1. **Tracking patterns.** Tracking patterns is intuitive for many people. Unlike anomalies, patterns are generally reliable, though they're by no means infallible. One of the most basic techniques in data mining is learning to recognize patterns in your data sets. This is usually a recognition of some aberration in your data happening at regular intervals, or an ebb and flow of a certain variable over time. [6]
2. **Classification.** Classification is a more complex data mining technique that forces you to collect various attributes together into discernable categories, which you can then use to draw further conclusions, or serve some function, [6]
3. **Association.** Association is related to tracking patterns, but is more specific to dependently linked variables. In this case, you'll look for specific events or attributes that are highly correlated with another event or attribute [6]
4. **Outlier detection.** Outlier detection is the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data. In many cases, simply recognizing the overarching pattern can't give you a clear understanding of your data set. You also need to be able to identify anomalies, or outliers in your data. [6]
5. **Clustering.** Clustering is very similar to classification, but involves grouping chunks of data together based on their similarities. [6]
6. **Regression.** Regression, used primarily as a form of planning and modeling, is used to identify the likelihood of a certain variable, given the presence of other variables. [6]
7. **Prediction.** Prediction is one of the most valuable data mining techniques, since it's used to project the types of data you'll see in the future. In many cases, just recognizing and understanding historical trends is enough to chart a somewhat accurate prediction of what will happen in the future. [6]

IV. Data Mining Applications:
Several data mining applications have been successfully in forced in diverse areas like health care, finance, retail, telecommunication, fraud detection and risk analysis etc. The ever increasing complexities in several fields and improvements in technology have posed new challenges to data mining; the several challenges include different data formats, data from disparate locations, advances in computation and networking resources, research and scientific fields, ever growing business challenges etc. [5] Data mining applications can be developed to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims. It can search for patterns that might indicate an attack by bioterrorists. Moreover, this system can be used for hospital infection control, or as an automated early warning system in the event of epidemics. [5]

V. KNOWLEDGE DISCOVERY DATABASE
The Data Mining and KDD often used interchangeably because Data mining is the key part of KDD process. The goal of the KDD process is to extract knowledge from data in the context of large data bases. It does this by using data mining methods(algorithms) to extract (identify) what is deemed knowledge, according to the specifications of measures and thresholds, using a database along with any required preprocessing, sub sampling, and transformations of that database. KDD field is concerned with the development of methods and techniques for making sense of data. At the core of the process is the application of specific data-mining methods for pattern discovery and extraction. [2]

Knowledge discovery in databases (KDD) is the process of discovering useful knowledge from a collection of data. This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results. [7]

KDD Techniques: Learning algorithms are an integral part of KDD. Learning techniques may be supervised or unsupervised. In general, supervised learning techniques enjoy a better success rate as defined in terms of usefulness of discovered knowledge. According to [9], learning algorithms are complex and generally considered the hardest part of any KDD technique. Machine discovery is one of the earliest fields that has contributed to KDD [10]. While machine discovery relies solely on an autonomous approach to information discovery, KDD typically combines automated approaches with human interaction to assure accurate, useful, and understandable results.

There are many different approaches that are classified as KDD techniques. There are quantitative approaches, such as the probabilistic and statistical approaches. There are approaches that utilize visualization techniques. There are classification approaches such as Bayesian classification, inductive logic, data cleaning/pattern discovery, and decision tree analysis. Other approaches include deviation and trend analysis, genetic algorithms, neural networks, and hybrid approaches that combine two or more techniques. [8] Because of the ways that these techniques can be used and combined, there is a lack of agreement on how these techniques should be categorize.

The usefulness of future applications of KDD is far-reaching. KDD may be used as a means of information retrieval; in the same manner that intelligent agents perform information retrieval on the web. New patterns or trends in data may be discovered using these techniques. KDD may also be used as a basis for the intelligent interfaces of tomorrow, by adding a knowledge discovery component to a database engine or by integrating KDD with spreadsheets and visualizations. [8]

KDD Applications: Major KDD application areas include marketing, fraud detection, telecommunication and manufacturing. Other applications of KDD in healthcare are many providers are migrating toward the use EHR store a large quantity of patient data on test results, medications, prior diagnoses, and other medical history. This is a valuable source of information that could be better used by employing KDD techniques. Several examples include identifying patients who should receive flu shots, enroll in a

disease management program and are not in compliance with a treatment plan. Moreover, historical data in EHR can help in management of chronic diseases and anticipating patient's future behavior on the given history. EHR stores spatial and demographic data which can help in public health management and planning. [5]

KDD includes multidisciplinary activities. This encompasses data storage and access, scaling algorithms to massive data sets and interpreting results. The data cleansing and data access process included in data warehousing facilitate the KDD process. Artificial intelligence also supports KDD by discovering empirical laws from experimentation and observations. The patterns recognized in the data must be valid on new data, and possess some degree of certainty. These patterns are considered new knowledge. Steps involved in the entire KDD process are:

1. Identify the goal of the KDD process from the customer's perspective.
2. Understand application domains involved and the knowledge that's required
3. Select a target data set or subset of data samples on which discovery is be performed.
4. Cleanse and preprocess data by deciding strategies to handle missing fields and alter the data as per the requirements.
5. Simplify the data sets by removing unwanted variables. Then, analyze useful features that can be used to represent the data, depending on the goal or task.
6. Match KDD goals with data mining methods to suggest hidden patterns.
7. Choose data mining algorithms to discover hidden patterns. This process includes deciding which models and parameters might be appropriate for the overall KDD process. [7]
8. Search for patterns of interest in a particular representational form, which include classification rules or trees, regression and clustering.
9. Interpret essential knowledge from the mined patterns.
10. Use the knowledge and incorporate it into another system for further action.
11. Document it and make reports for interested parties. [7]

VI. CONCLUSION

In real-time information technology has generated and used large amount of databases and stored huge data in various areas. [3] An overview of knowledge discovery database and data mining techniques has provided an extensive study on data mining techniques. Data mining has the most important

and talented features of interdisciplinary developments in Information technology.

REFERENCES

[1] https://www.researchgate.net/journal/1384-5810_Data_Mining_and_Knowledge_Discovery

[2] Priyadharsini.C and Dr. Antony Selvadoss Thanamani," An Overview of Knowledge Discovery Database and Data mining Techniques". International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

[3] Mrs. Bharati M. Ramageri, "Data Mining Techniques and Applications," Indian Journal of Computer Science and Engineering, Vol. 1 No. 4, pp. 301-305

[4] Kalyani M Raval, "Data Mining Techniques," International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2 Issue 10,pp.439-442.

[5] Mohit Saini, (2018)." DATA MINING TREND IN PAST, CURRENT AND FUTURE".

[6] https://www.datasciencecentral.com/profiles/blogs/the-7-most-important-data-mining-techniques

[7] https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd

[8] Peggy Wright, "Knowledge Discovery In Databases: Tools and Techniques". http://www.mariapinto.es/ciberabstracts/Articulos/Knowledge%20Discovery.htm

[9] Brachman, R. J., and Anand, T. The Process Of Knowledge Discovery In Databases: A Human-Centered Approach. In *Advances In Knowledge Discovery And Data Mining*, eds. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, AAAI Press/The MIT Press, Menlo Park, CA. 1996, pp. 37-57.

[10] Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P. From Data Mining To Knowledge Discovery: An Overview. In *Advances In Knowledge Discovery And Data Mining*, eds. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, AAAI Press/The MIT Press, Menlo Park, CA., 1996, pp. 1-34.

[11] Sasisekharan, R., Seshadri, V., and Weiss, S. M. Data Mining And Forecasting In Large-Scale Telecommunication Networks. *IEEE Expert: Intelligent Systems & Their Applications* 11, 1 (Feb. 1996), 37-43.