Knowledge Discovery and Data Mining Review Papers

Zenawi Haileslassie Berhe (MTech) *[1], Associate professor Vinod Kumar *[2]
*Department of Computer Engineering, Delhi Technological University, Delhi, India *[1]*
*Department of Computer Engineering, Delhi Technological University, Delhi, India *[2]*

Abstract

The main objective of this paper is on challenges in data mining and knowledge discovery and different types of knowledge discovery in database techniques like visualization, statistical, different types of machine learning, and different types of databases are highly used in knowledge discovery and data mining. Knowledge discovery and data mining also mainly focus on datamining technology, data mining tasks, different types of knowledge discovery in database or KDD process are the main thigs focused on this paper [1][2][3][4][5][6].

Keywords - *data mining tasks, knowledge discovery, Large Databases, KDD process*

I. INTRODUCTION

Knowledge discovery and data mining is most of the time focused on combinations of different necessary data points. And knowledge discovery and data mining satisfy the new things that is generated in our daily life of different computational techniques and different toolsused in different information's during extracting and assisting of different data mining points. These kinds of information's are also known as or considered as knowledge. On this paper mainly focused on how these different types of techniques are used and how the different types of tools and subject of the knowledge discovery in dataare knowledge discovery and data mining are used in processing different data mining actives and also different sets how to generate and how to extract with the help of data mining work [1].

In data mining and knowledge discovery there are different types of approaches and these different types of approaches can be categorized in to four main categories base on their operations of data mining techniques that are included in different techniques used in knowledge discovery and data mining as follows: -

The first approach is predictive model creation is one the supervised machine learning techniques used in different analyzing and also discover different producers' step by step in database.

The second type of approach in data mining and knowledge discovery is segmentation. Segmentation is also used in different types of techniques most of the time.

The third type of approaches is also detection deviation this is used most of the time use in different statistical techniques [1].

II. THE MAIN DIFFERENCE BETWEEN DATA MINING AND KNOWLEDGE DISCOVERY IN DATABASE AND ITS BASIC CONCEPTS

The main difference between data mining and knowledge discovery in database is data mining is most of the time used to analyze different types of executive summary and different types of important information's how to analyze and how to identify with different categories of data mining and knowledge discovery and their angles of the different techniques and their relationships [3].

Basically, knowledge discovery and data mining are also finding with different correlations and regressions data is extracted from huge database.

A. The main tasks of data mining are

The main tasks of data miningare focused with different kinds of techniques used and generated data from data mining and knowledge discovery in database (KDD) or huge amount of data.

And another basic task of data mining is also by classifying theses different types of techniques and tools how to organize them to generate or extract data from large amount of data in the data base [5].

1. Summarization

Summarization is the deliberation or speculation of information. A set of assignment pertinent information is condensed also, disconnected. This outcomes in a littler set which gives a general outline of the information, for the most part with total data.

2. Classification

Classification gathers a capacity or model which decides the class of an object dependent on its

attributes. A lot of articles are given as the preparation set. In it, each item is spoken to by a vector of traits alongside its class. a grouping capacity or model is developed by breaking down the connection between the characteristics and the classes of the items in the preparation set. This capacity or model would then be able to order future items. This causes us build up a superior comprehension of the classes of the protests in the database.

3. Clustering

Bunching recognizes classes additionally called bunches or gatherings for a lot of objects whose classes are obscure. The objects are clustered to the point that the interclass similitudes are augmented and the interclass similitudes are limited. This is done based on certain criteria characterized on the properties of the articles. When the bunches are chosen, the objects are marked with their relating bunches. The regular highlights for articles in a bunch are condensed to shape the class depiction.

4. Trend investigation

Time arrangement information are records aggregated after some time. For instance, an organization's deals, a client's Mastercard exchanges and stock costs are untouched arrangement information. Such information can be seen as articles with a property time. The items are previews of substances with qualities that change after some time. Finding the examples and regularities in the information advancements along the component of time can be entrancing.

B. Some of the steps of knowledge discovery in data base process

The knowledge discovery in data base process is the most type of interactive methods in different persuaders.

Knowledge discovery in data base (KDD)is divided in to different steps as follows: -

The first type of knowledge discovery in database (KDD) process is data selection and the second of knowledge discovery in database (KDD) process is data cleaning.

The third type of knowledge discovery in database (KDD) process is data transformation and the fourth type of knowledge discovery in database (KDD) process is pattern searching (data mining), and the last type of knowledge discovery in database (KDD) process is Finding presentation and also finding interpretation and finding evaluation[2][4].

Knowledge (Information) Discovery in Databases is the non-insignificant procedure of recognizing legitimate, novel, possibly helpful, and at last reasonable examples in information. Here information is a lot of actualities (e.g., cases in a database) and example is an articulation in a few languages portraying a subset of the information or a demonstrate pertinent to that subset. Removing a design assigns fitting a model to information, discovering structure from information, or all in all any abnormal state depiction of a lot of information. The term process infers that KDD is involved numerous means, which include information readiness, scan for examples, information assessment, and refinement, all rehashed in numerous emphases.

The found examples ought to be substantial on new information with some level of sureness. We likewise need examples to be novel (in any event to the framework, and ideally to the client) and conceivably helpful. At long last, the examples ought to be justifiable, in the event that not promptly, at that point after some post preparing.

The KDD procedure is intelligent and iterative, including various strides with numerous choices being made by the client. Here we extensively diagram a portion of its fundamental advances:

- ✓ Learning the application space:
  - Relevant earlier information and objectives of application
- ✓ Creating an objective informational index: information choice
- ✓ Data cleaning and preprocessing
- ✓ Data decrease and projection:
  - Find helpful highlights, dimensionality/variable decrease and invariant portrayal.
- ✓ Choosing the mining algorithm(s)
- ✓ Data mining: scan for examples of intrigue
- ✓ Interpretation: investigation of results.
  - Visualization, change, evacuating repetitive designs.
- ✓ Use of found learning are some of knowledge discovery procedures.

C. Some of Data mining challenges are

The most commonly challenges that are most of the time happen in different researches and also in the domains of knowledge discovery in data base is most of the time happened in our daily activities[1][3].

Basically, the most commonly challenges happened in data mining are the following: -

The first and basic and commonly happened challenge is normalization.

The second challenge in data mining is estimation missing value and the third challenge is detection outlier.

III. DATA MINING APPLICATIONS

The main applications of data mining are in short and précised form is the following things: -

Some of the applications of data mining is advertising, different bioinformatics, in different Customer Relationship Management (CRM), database Marketing, E-Commerce, health Care, investment/Securities, different manufacturing, Process Control, different sports activities and Entertainment, telecommunications, different web etc.… [6].

Information mining systems have been connected effectively in numerous territories from business to science to help the following additional things: -

A. Business applications

Many associations presently utilize information mining as a mystery weapon to keep or pick up a focused edge. Information mining has been utilized in database promoting, retail information investigation, stock determination, credit endorsement, and so forth.

B. Science applications

Data mining systems have been utilized in cosmology, sub-atomic science, prescription, geography and some more.

C. Other applications

Data mining systems have likewise been utilized in medicinal services the executives, charge extortion location, and illegal tax avoidance observing and even games.

IV. CONCLUSION

This paper is focused on data mining and knowledge discovery in database (KDD) and what are the challenge, applications and what are the tasks and process of knowledge discovery and data mining in short and also what are the different types of techniques in data mining, knowledge discovery and its basic concepts[4][5].

REFERENCES

[1] Mykhaylo Lobur1, Yuri Stekh2, Vitalij Artsibasov3" challenges in knowledge discovery and data mining in datasets",IEEE Expert, Vol. 9, No. 1,1994, pp. 60–66.

[2] Qi Luo1, 2 "advancing knowledge discovery and data mining", 0-7695-3090-7/08 $25.00 © 2008 IEEE

[3] M.Lobur, Yu. Stekh, A.Kernytskyy, Faisal M.E. Sardieh" Some Trends in Knowledge Discovery and Data Mining"MEMSTECH'2008, May 21-24, 2008, Polyana, UKRAINE.

[4] M-Tahar Kechadi1, and Ilias K. Savvas2 1University College Dublin, Ireland, 2T.E.I. of Larissa, Greece, tahar.kechadi@ucd.ie, savvas@teilar.gr" Cooperative Knowledge Discovery & Data Mining CKDD" 2012 IEEE 21st International WETICE.

[5] M.Lobur, Yu. Stekh, A.Kernytskyy, Faisal M.E. Sardieh" Some Trends in Knowledge Discovery and Data Mining"MEMSTECH'2008, May 21-24, 2008, Polyana, UKRAINE.

[6] Manish Marwah1, Ratnesh Sharma1, Naren Ramakrishnan2, Cullen Bash1, Chandrakant Patel1 1HP Labs, Palo Alto, CA 2Virginia Tech, Blacksburg, VA"Knowledge Discovery and Data Mining for Enhanced Sustainability of Physical Ecosystems"ITHERM'08, June 2008, Orlando, FL.