International Conference on Latest Innovation in Engineering Science and Management
Buddha Institute of Technology, GIDA, Gorakhpur (UP)
30th - 31st March 2019 www.conferenceworld.in ISBN : 978-93-87793-80-4

Data Discovery on the basis of knowledge

Soni Mishra[1],Pragati Srivastava[2],Nisha Yadav[3],Soni Singh[4]
Buddha Institute of Technology
Gida,Gorakhpur UP INDIA- 273209
[1]*Department of Information technology,Abdul Kalam Technical University Lucknow;*
[2]*Department of Information technology,Abdul Kalam Technical University Lucknow;*
[3]*Department of Information technology,Abdul Kalam Technical University Lucknow;*
[4]*Department of Information technology,Abdul Kalam Technical University Lucknow;*

ABSTRACT :
The data discovery on the basis of knowledge is very dynamic research and developlopement area.the data discovery is a procces for extract data in the database or any field of data on the basis of knowledge.It requires well-define foundation, This type of discovery is well understood and popularized throughout the community.This survey is represent historical overview,description and future plan concerning a standard for data discovery on the basis of knowledge.It leads the several process model,and discusses their applications to academic and industrial problems.The main goal of this review is the consolidation of the research in this area.This survey is also proposes to enhence property of the data.

I. INTRODUCTION

Data Discovery on the basis of knowledge is the most desirable end-product of computing. Finding new method and enhancing our knowledge in any data set field.It provide a greater long-range value for optimization production processes or inventories . it is most difficult for computing challenges to do well Gio Wiederhold (1996). it support decisions and policies made by scientists and businesses (Fayyad et al., 1996c).

Current technological is permits the storage and access of large amounts of data at virtually without cost .The true value is not in storing the data, but it has ability to extract useful and to data find interesting trends and correlations, through the use of statistical analysis and inference.

Before any attempt can be made to perform the extraction of this useful data on the basis of knowledge, an overall approach that describes how to extract the data on the basis of Knowledge . Data Discovery is provide a road map to follow while planning and carrying out the projects . So it gives the results time and cost savings, and in a better understanding of the data . Therefore, the focus of this paper is not on describing the methods that can be used to extract the data on the basis of knowledge, but rather on discussing the methodology that supports the process that leads to finding this knowledge.

This procces is supports the business plan and solve the industerial problem and provide the usefull data for extuction of any plan .

This survey is organized as follows . First, basic definitions concerning the Data Discovery on the basis of knowledge , motivation and a historical overview are provided in section . 2 In the Next section provide overview of current technology and data storing procces . 3 In this section discribre the methodlogy supports and data extraction . 4 Finally , in this section discussed future plan for any data set area and

conclusion.

II. PROCESS MODEL OF DATA DISCOERY ON BASIC OF KNOWLEDGE:

2.1. Terminology:

There is a common confusion in understanding the terms of Data Discovery on the basis of knowledge , the meanings of these terms explained with the help of knowledge discovery database and data mining process . Data Discovery is also known under many other names, including knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing (Fayyad et al., 1996c) . Data Discovery is a process that seeks new knowledge about an application domain . The aim at completion of a particular discovery task, and accomplished by the application of a discovery method (Klosgen & Zytkow, 1996).

2.2. Motivation:

The main motivation factor to formally structure of the Data Discovery on the basis of knowledge as a process results from anobservation of problems associated with a blind application of DM methods to input data . Such activity, called 'data dredging' in the statistical literature, can lead to discovery of meaningless knowledge (Fayyad et al., 1996c).

Lastly, there is a widely recognized need for the standardization of data discovery on the basis of knowledge this process is provide a unified view on existing process descriptions and to allow an appropriate usage of technology to solve current business problem.

III . Data Extraction on the Basic of Knowledge:

This process is most popular for extraction of usefull data .This process is known as Knowledge Discovery database. It refers to the nontrival extraction of previously unknown and potentially useul information from data in database . Data Discovery is a part of Knowledge Discovery process . This mode can be applied to specific scenarios.

such as: Preparing Data ,Exploring Data ,Building Models, Exploring and Validating Models.

3.1. Forecasting: Estimating sales, reprisent server loads or server downtime.

3.2. Risk and probability: Choosing the best customers for targeted , determining the probablem break-even point for risk scenarios, assigning probabilities to diagnoses or other outcomes.

3.3. Recommendations: Determining which products are likely to be sold together, generating recommendations.

3.4. Finding sequences: Analyzing customer selections in a shopping cart, predicting next likely events.

3.5. Grouping: Separating customers or events into cluster of related item.

4. History :

in the early 1990s,when the KDD process termd was first coined (Piatetsky-Shapro and Frawley in 1991),there was developed in Rush through the Data mining algorithm that are capable to solving all the problem of searching in data . The concept of a data dicovery process model is similar to knowledge discovery database originally discussed during the first workshop on KDD in 1989 (Piatesky-Shapiro, 1991). The overall process framework (Zytow & Baker, 1991; Klosgen, 1992; Piatesky-Shapiro & Matheus, 1992; Ziarko et al., 1993; Simoudis et al., 1994). Such systems were intended for expert users who had understanding of DM techniques, the underlying data, and the knowledge sought.

FIGURE OF DATA EXTRACTION:

5.APPPLICATION :

5.1. Applications of KDDM models:

To complement the description of existing data discovery models, their applications to a variety of , in both industrial and research communities. Applications are grouped by the model used. The nine-step model by Fayyad et al. is the most cited model in the professional literature to date. It has been

ISBN : 978-93-87793-80-4

incorporated into an industrial DM software system called Mining (Brunk et al., 1997), and applied in a number of KDDM projects (all projects, except the last, are predominantly research oriented) : research and industrial domains are briefly summarize data .This is mainly due to the fact that research users typically know the data much better than industrial users; they have better understanding of novel technologies, and are better trained to organize intuitions int computerize procedures (Fayyad et al., 1996f). This observation only corroborates the need for standard process models to support planning and execution of KDDM projects in industrial communities.

6. TABLE:

Table 6.1:

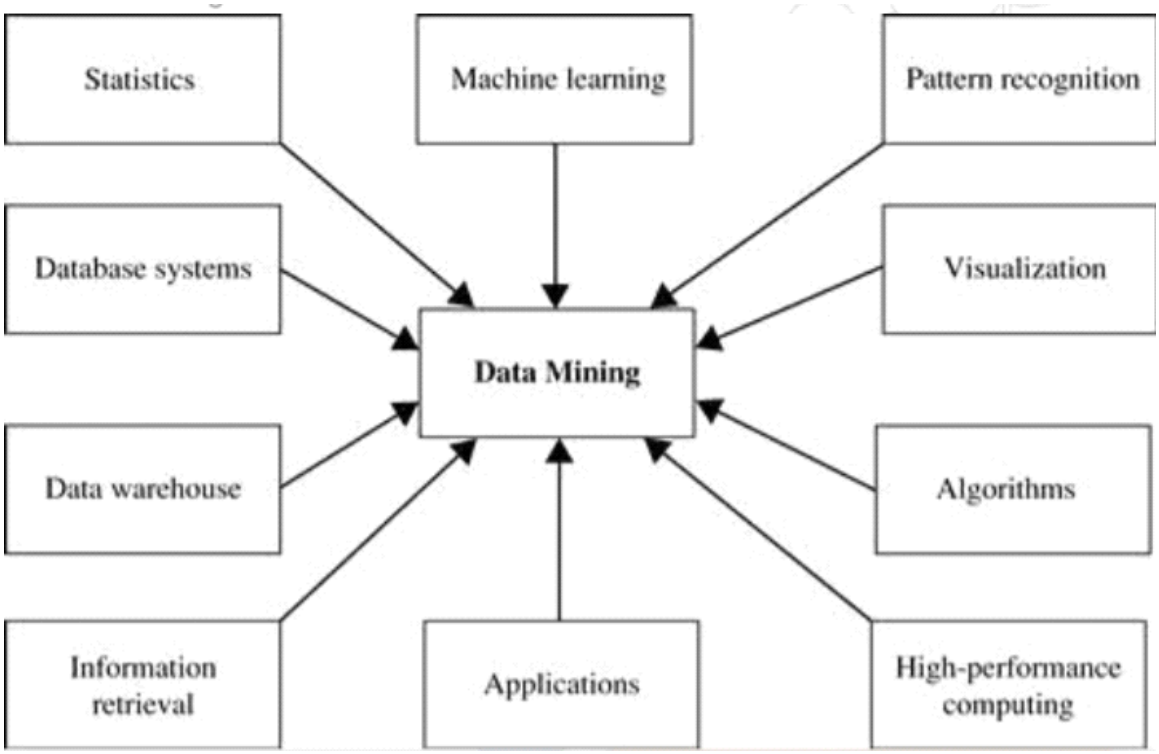
Fayyad et al.'s (1996) KDD Process-centric framework	Han & Kamber's (2001) Proposed Methodology	The CRISP-DM Methodology	Our Method
[1] Learning the application domain	(1) Problem analysis	(1) Business understanding	(1) Learn domain of practice ,and decision problem (2) Structure and formalize relevant domain prior knowledge
[2-4]Creating the target dataset; Cleaning and pre-processing; data reduction and data selection	(2) Data preparation (3) Data exploration	(2) Data understanding (3) Data preparation	(2) Data understanding (3) Data preparation: incorporate additional formalized prior knowledge in machine-readable format
[5-7] Choosing the data mining technique; Choosing the data mining algorithm; Data mining	(4) Pattern generation (data mining)	(4) Modeling	(4) Modeling
[8] Interpreting the results	(5) Pattern monitoring	(5) Evaluation	(5) Evaluation
[9] Using the discovered knowledge	(6) Pattern deployment.	(6) Deployment.	(6) Deployment.

ISBN : 978-93-87793-80-4

Table 6.2:

Step	Business Question	Enabling Technologies	Product Providers	Characteristics
Data Collection (1960s)	"What was my total revenue in the last five years?"	Computers, tapes, disks	IBM, CDC	Retrospective, static data delivery
Data Access (1980s)	"What were unit sales in New England last March?"	Relational databases (RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Retrospective, dynamic data delivery at record level
Data Warehousing & Decision Support (1990s)	"What were unit sales in New England last March? Drill down to Boston."	On-line analytic processing (OLAP), multidimensional databases, data warehouses	Pilot, Comshare, Arbor, Cognos, Microstrategy	Retrospective, dynamic data delivery at multiple levels
Data Mining (Emerging Today)	"What's likely to happen to Boston unit sales next month? Why?"	Advanced algorithms, multiprocessor computers, massive databases	Pilot, Lockheed, IBM, SGI, numerous startups (nascent industry)	Prospective, proactive information delivery

7.Figure of Data extraction:



8. Conclusions and summary :

The Data Discovery on basis of knowledge is the verge of processing one of the most successful business technologies. Which stands in the way to the success is the in accessibility of the related applications to broad scientific and industrial communities. This shortcoming can be overcome only by moving beyond its algorithm-centric roots (Apps, 2000). The challenge for the 21st century data miners is to developand popularize widely accepted standards that, if adopted, will stimulate major industry growth and interest (Piatesky-Shapiro, 1999a) . The fuel for the growth is the strong economic and social need for solutions provided by the KDDM community (Fayyad et al., 1996c). This survey has provided an overview of the state-of-the-art in developing one of the most important standards, the KDDM process model .On the main goal of this survey has been to consolidate research in this area , to inform users has different models and how to select the appropriate method, and to develop improvedthe method that are based on previous experiences. It promote development and delivery of solutions that use business language, and traditional language of algorithms, matrices, criterions, complexities, and the it gives the result in a greater exposure and acceptance of the KDDM industry. This, will be a significant factor in pushing the industry beyond the edge, and into the mainstream.

10. Acknowledgements :

The Discovery of data is an itractive process .Once the discovered knowledge is presented to the user ,the

ISBN : 978-93-87793-80-4

evaluation measure can be enhenched ,the mining can be refind new data,more appopriate result.
The data mining derives its name from the similarities between searching for valuable information in large database and mining for valuable .Other similar turms referring to data mining are : Data removing , knowledge extraction and pattren discovery.

REFERENCES:

BOOKS:

[1] *Gajendra Sharma ,Data mining data warehouse and OLAP, 2007-2008.*

[2] *Alex Berson ,Stephen J.Smith,Data warehousing Data miming,&OLAP,1997 by McGraw-Hill Companies*

WEBSITE:

[1]http://www.tutorialspoint.com

[2] http://www.exastax.com

[3]http://machinelearningnastery.com

THESIS:

[1]http://www.techscarks.com.in

[2]http://www.writenythesis.org