An Transpire Trends Survey on Knowledge Innovation of Data Mining and Web Mining

Dr. P. Sengottuvelan

Associate Professor
Department of Computer Science
Periyar University PG Ext. Centre
Dharmapuri-636705

N.Revathi

Ph.D Research Scholar
Department of Computer Science
Periyar University PG Ext. Centre
Dharmapuri-636705

Abstract: Data mining is an interesting growing incorporative of sub domain. It is knowledge management of Database system. The Data mining is an analyzing process of discover hidden information and pattern from large scale of data. Data removal influences on various field as organizations such as Statistics. The Web mining is the perception of data mining technique and application of mining for facts illustration out from data. The data is collected from web server and huge scale data for finding access prototype because, unprocessed data which is collected from the web server is incomplete. Aspire of this paper is to recognize the preprocessing of usage data and also the discovery of Patterns and their analysis.

Keywords: Knowledge discovery, Predictive Mining, Web Usage Mining, Data mining, Pattern discovery.

I. INTRODUCTION

The Data mining concept related to the making information for decision making and business is very poor even though data storage grows exponentially. Data mining concept is also known as Information exploration of Database system.

Emerging process that analyses the enormous amount of data and extract the exact knowledge from the data. Web mining is used to retrieve the information then extract and evaluate data for information discovery from documents on Web. Web mining consists of Web contented, Web structure analysis and Web usage mining[3]. Web Content Mining deals with the discovery of information which is useful from the web data or documents. Web Structure mines the hyperlinks structure within the web page. Data mine is correlated on different industries like retail, finance, health care, aerospace, educationsn etc. Knowledge is extracted from the historical data by applying pattern recognition, statistical and mathematical techniques that results in the knowledge in the form of facts, trends, associations, patterns, anomalies and exceptions.

II. LITERATURE SURVEY

The main problem given is to get a dataset which is reliable for mining. Hence the data should be retreated and accessing behavior of users is to be constructed as transactions. The reliability of a transaction is an important. By using Cookies or authentication mechanism, users are identified. But users are not attracted by these types of sites due to privacy concerns. The two heuristics are mentioned for the acknowledgement of requests to different visitors. It has undergone a range of steps such as data attack, user recognition, session classification and clustering. Also the communication which show the user's behavior are construct exactly in preprocessing step by manipulative the Reference Lengths of user access by means of byte rate. Also by using Maximal Forward Reference [4] and Reference Length algorithm Time Window concept is combined to find pages carries contents. By using Web clustering the objects a variety of types can be clustered into different groups for various purposes.

Uma Maheswari, Dr. P. Sumathi to given that for discovering patterns, the sessions are to be constructed efficiently and also reviews existing work done in the preprocessing stage[1]. The results of mining can be used to improve the design of the website and increase satisfaction which helps in various applications. The raw log

files contain unnecessary details like images which are accessed, breakdown entries etc., which will affect the exactness of pattern innovation Phase and analysis Phase. So the preprocessing stage is significant work in removal to make efficient pattern analysis. To get accurate mining results the session details of users are to be known. Thair Nu Phyu to the survey was performed on a selection of web usage methods used in preprocessing proposed by research community[2]. The preprocessing stages like session identification and path completion on which more concentration is done and this paper also have presented various works done by different researchers.

Navin Kumar Tyagi et al[3]., to give a clear understanding process of the data preparation and discovery of patterns. This paper provides a clear idea about the pattern discovery in the web usage mining process. The various algorithms are proposed in this paper for data preprocessing like data Preparation, user identification and session identification.

III. DATA MINING PERFORMANCE

The data removal methods to take out patterns from data. Each method has different aim, which decides the outcome of the KDD process entirely [8].

A. Predictive Mining:

Supervised learning task where the unknown value of a class or future values of interest is predict from the existing data. It can also validate a newly make-believe hypothesis.

1). Classification: It is results of categorization model termed as classifiers the data as classes and concepts.

2). Regression: Data mining occupation is used to predict the missing or unavailable arithmetic data values by mapping the data into a function. Linear regression and Multi- Linear regression are some of the regression techniques.

B. Descriptive mining:

It is a task of abbreviation the data and its benefits as patterns using for data mining and data aggregation methods.

1) Clustering: Clustering is a task of grouping data of similar characteristics into a cluster while the different data may group into different respective clusters. Search for the cluster is a unsupervised learning.

2) Association rule mining: Association rule mining unwraps the patterns that occur frequently among the data set. It focus in extracting associations, correlations, frequent sequence, frequent item set and frequent patterns with interestingness among the data set in the data repositories.

3) Summarization: Summarization is the process of reducing the huge volume of data in a meaningful and intelligent fashion with important and relevant features. Summarization techniques like tabulation of the mean and the standard deviations are often implied to analyses and visualize the data,

C. R – Programming: It is developed from C and FORTRAN. It's a freeware that provide software programming language for graphics. Data miners to develop statistical software and data analysis.

IV. WEB USAGE MINING

Web usage mining is the application of data mining techniques on large web log repositories to discover knowledge which is useful about behavioural pattern of user and also website usage statistics that can be used for various website design tasks [5]. Data Collection: The data in log is collected from sources like server side, client side and proxy servers and so on. Data Preprocessing: This is done on raw data which present in log file wrapping up of data cleaning, user identification and session identification. Pattern discovery:

- Data Collection
- Data Preprocessing
- Pattern Analysis
- Pattern Discovery

1. Data Collection:

The data collection step includes various data sources. Mining Process the Primitive source of data in web removal is the log at server. There are some additional data source are also use for some user and some application which includes log on client side and Proxy side log [3]. In Log at client side, usage data can be tracked.

2. Data Preprocessing:

The information available in the web is Varied and unstructured. Therefore, the preprocessing phase is a required for discovering patterns. The purpose of this is to transform the raw data into a group of user profiles. Data preprocessing is important and this led to various algorithms and Socratic techniques for it such as Data

Cleaning, User and Session Classifications etc.

3. Pattern Discovery:
Once transactions of user have been identified, techniques of data mining are performed for pattern discovery in web usage mining process. These methods represent the ways that often appear in the data mining study such as discovery of association rules and sequential patterns and clustering and classification etc. Classification is a supervised learning process. In this, the data item mapped into one of several predefined classes.

4. Pattern Analysis:
The last stage of web usage mining Process is Pattern Analysis. The patterns which re mined are not suitable for interpretations. So it is important to sort out patterns or rules which are not interesting from the set found in the pattern discovery phase. The tools are provided to help the transformation of information into knowledge.

V. APPLICATIONS OF DATA MINING
Data Mining is used in many domains in constant basis. Some of these organizations include retail stores, hospitals, banks, and insurance companies. Many domains like health care, finance insurance, retail stores combines the data mining with as statistics, pattern recognition, and other important tools to perform data analytics. Data mining is used primarily for decision making.

A. Medicare and health care:
Using data mining methods, it is able to find the correlation between the diseases, to analyze the effectiveness of the treatments given, to identify the new drugs to analyze the market of the drugs and etc.

B. Education:
Educational Data Mining is a blooming field which knowledge from educational Environment data. The goals of EDM are identified as predicting students' learning behavior, emotions, and skills. This study improves the educating methods by understanding the ward and to take accurate decisions respectively.

C. Market Basket Analysis:
This technique that uses association rule mining to understand the purchasing behavior of the customer. It also allows the seller to understand his business, customer's needs and to make profitable change accordingly.

D. Financial Banking:
Data mining can subscribe to solving business harms in banking and finance by sentence patterns, causalities, and correlation in business information and market prices. The managers may find this information for better segmenting, targeting, acquiring, retaining and maintaining a profitable customer.

E. Research Analysis:
Data mining is very useful in data pre-processing and integration of databases. Data mining allows the supporter to identify co-occurring sequence and the organization between any activities. Data visualization and visual data mining help the researcher with a clear view of the data [6].

F. Fraud Detection:
Huge amount of dollars is being lost because of fraud detection. Traditional methods are time consuming and complex. Data mining aids in providing significant pattern and turning data into information. Valid and useful information is called as knowledge. The results are categorized into fraudulent or non-fraudulent. In calculation to this it plays a very vital role in Bio- Informatics, Criminal Investigation, Corporate Surveillance, Insurance, Agriculture, Customer Segmentation, Lie Detection Intrusion Detection, Manufacture engineering and etc.

VI. CONCLUSION
Finally the process of KDD and relevance of data mining in various sectors is discussed. The data mining functionalities predictive mining (classification, regression, prediction and decision trees) and descriptive mining (clustering, association, and summarization) is also summarizing. The necessary of data mining in commercial, educational, medical, scientific fields are highlighted. The Data Preprocessing is very necessary. Due to this all irrelevant entries in the dataset are deleted. And only remains the important data which is used for the next phases in the web usage taking out process such as mold detection and prototype analysis.

REFERENCES

[1]. Uma Maheswari, Dr. P.Sumathi, A New Clustering and Preprocessing for Web Log Mining, World Congress on Computingand Communication Technologies, IEEE, 2014

[2]. Theint Theint Aye, Web Log Cleaning for Mining of Web Usage Patterns, International conference on Computer Research and development, 2011, IEEE, Volume 2.

[3]. Navin Kumar Tyagi, A.K. Solanki, Sanjay Tyagi, An Algorithmic Approach to Data Preprocessing in Web Usage Mining http://thenewstack.io/six-of-the-best-open-source-data-mining-tools/

[4]. Thair Nu Phyu , Survey of Classification Techniques in Data Mining ,Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I,IMECS 2009, March 18 - 20, 2009, Hong Kong,ISBN: 978-988-17012-2-0 .

[5]. Lior Rokach, Oded Maimon, clustering methods, data mining and knowledge discovery handbook, Department of Industrial Engineering,Tel-Aviv University.

[6]. Sarat M. Kocherlakota, Christopher G. Healey,Summarization Techniques for Visualization of Large Multidimensional Datasets,Technical Report TR-2005-35Knowledge Discovery Lab,Department of Computer Science, North Carolina State University,Raleigh, NC 27695-8207.

[7]. T.Karthikeyan,N. Ravikumar A Survey on Association Rule Mining, International Journal of Advanced Research in Computer and Communication Engineering,Vol. 3, ISSN (Print) : 2319 5940,ISSN (Online) : 22781021, Issue 1,January 2014, http://bigdata-madesimple.com/14-useful-applications-of-data-minin.

[8]. Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems).