E-Commerce With Backbone Of Data Mining

I. Kali Pradeep
Assistant Professor
VIT ,Bhimavaram

I. Bhagyasri
Student
SCET,Narasapur

P. Praneetha
Assistant Professor
SCET, Narasapur

Abstract

E-commerce is changing the face of shopping day by day. E-Commerce came into existence few years ago and data mining has become a backbone for E-Commerce. Data mining in E-commerce is used to extract previously unknown information from huge amount of data by using different criteria. This data include different buying patterns and many interesting relationships among data. The paper describes an overview of interesting pattern analysis recommender systems and web usage mining related to E-commerce. And also describes the challenges faced by data mining in this case.

1. Introduction:

E-commerce has made life easier by enabling customers to buy goods of their choice from home. It enabled small vendors to sell their products online without much of advertising. But there are many background challenges faced by e-commerce companies to attract customers. In today's world, plenty of data is available with the E-commerce companies. This data is to be organized by using database tools and data warehouse. Data Mining provides many powerful techniques to convert this raw data into information which is used by decision makers to take path changing decisions. The hidden patterns have the mass power to change the shape of the e-commerce organizations.

Data mining use many algorithms for generating patterns. These patterns are help for detecting navigation style of the users, altering the structure of websites, dynamical content display to user and many more. Many data mining techniques like association analysis characterization, classification, cluster analysis, and outlier analysis are used for making E-commerce data more liable.

Data mining has many more challenges to deal with E-commerce data. This data may be abundant and it may contain structured, semi structured and unstructured data. Moreover, the data mining algorithms should be incremental, interactive, and should incorporate real time data.

The different functionalities of data mining like association analysis, cluster analysis, classification techniques, Outlier analysis and correlation analysis are used in every step of knowledge discovery process to discover the secret behind the good e-commerce websites

Many researchers contributed their work towards E-commerce in term of data mining. But, still E-commerce is not fully habituated in developing countries. And, E-Commerce is still limited to only little class of people. This paper presents a review on identifying interesting patterns, the recommender systems for e- commerce and web usage mining which can help E-Commerce websites to Enable more people to have a global Shopping experience.

2. Identifying interesting patterns:

Thousands of products are sold by the E-commerce companies day by day. And, every customer likes to have one to one marketing strategy based on his likes and dislikes. Many new techniques like dynamic web content presentations, purchase recommendations and targeting advertising are used in this context. Individual profile construction by association rule based discovery plays a key role in E-commerce websites. Apriori algorithms may generate many patterns, and manual inspection is not possible [1]. In many cases simple rules such as buys (bread) -> buys(butter) will not be sufficient to extract quality patterns. Some pruning techniques such as Incorporation of additional constraints, Redundancy reduction, Visualization, Organization and summarization, Rule grouping and clustering are used to generate useful patterns. By using these techniques uninteresting itemsets removal, alternate itemsets generation, itemsets comparison, are done dynamically.

3. Recommender systems in E-commerce:

Recommender systems help in presenting information such as recommended products or services to user based on his/her interests by applying knowledge discovery techniques. The information available with e-commerce companies are In terms of terabytes. And recommender systems are very important tool to deal with this data. There are many

recommender systems available some of which are highlighted here .The traditional recommenders are very simple recommenders which are still used by many companies. The purchase and navigation patterns are discovered using simple measures like support and confidence. Consider two sets of elements A and B, such that both A and B belongs to P (set of samples). A product is recommended with another product if they satisfy minimum support and minimum confidence. Where, support is represented by number of transactions containing both A and B to the total number of transactions. And, Confidence is represented by number of transaction containing both A and B to the number of transactions containing B.

A predictive event notification systems has been used [3] which is based on association analysis, clustering analysis, fuzzy computing and interval computing. It has mainly four components namely event manager, event channel manager, registry and proxy managers. This system not only triggers the customer on events but also predict event and event classes that are likely to be viewed by the customer.

Niu et al [9] present a method to build customer profiles in e-commerce settings, based on product hierarchy for more effective personalization. They divide each customer profile into three parts: basic profile learned from customer demographic data; preference profile learned from behavioral data, and rule profile mainly referring to association rules. Based on customer profiles, the authors generate two kinds of recommendations, which are interest recommendation and association recommendation. They also propose a special data structure called profile tree for effective searching and matching.
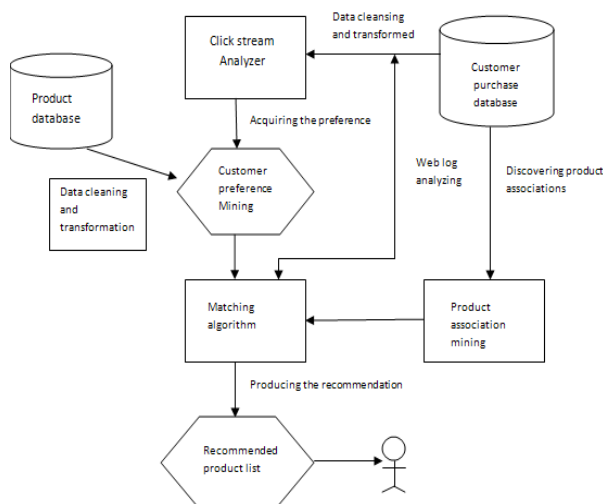

Fig 1: Recommender system for e-commerce

Location based advertising recommender for mobile users is the recommender systems that can be applied to recommend the advertisements of location-based services to mobile users LARMU is based on a collaborative filtering (CF) algorithm[7]. In general, CF works by building a user-item matrix that represents the satisfaction levels of users for items. CF in LARMU reflects other additional information, location, time, and user's needs type to generate personalized recommendations for mobile users. The E-commerce data from product database and customer purchase database along with weblog database is cleaned and transformed. This data is given as input to the matching algorithm where product associations are preserved. And at the last the recommender product list is shown to the user.

4. Web usage Mining

Web mining is the process of using data mining for extracting useful patterns from the web. These extracted patterns are used to improve the availability of information in the websites and the way those pieces of information are introduced to the website user, and to improve data retrieval and the quality of automatic search of information resources available in the web. Web mining can be divided into three major categories: web usage mining, web content mining, and web structure mining.

When weblog data is used as input for data mining engine in order to extract unknown and useful information about user access patterns, it is known as web usage mining. Web mining uses the data generated from user's sessions or behaviors. The web usage data includes data from web server access log, proxy server logs, browser logs, user profiles, registration data, cookies, and user queries. Web usage mining tries to predict user behavior while user interacts with the web and learns user navigation patterns. The learned knowledge could then be used for different applications such as website personalization, business intelligence, usage characterization and adaptive websites. The Web usage mining process consists of three phases: data preprocessing, pattern discovery, and pattern analysis

4.1 Data preprocessing

The conversion of user information into the format of data abstraction, which is an essential part of pattern discovery, is preprocessing. According to the preprocessing data, it is categorized into three parts: Usage Preprocessing, Content Preprocessing and Structure Preprocessing.

Usage preprocessing is the one of the difficult task in web usage mining. They gather data from IP address, agents and server side click streams, because of the nature of data, always the data is incomplete.
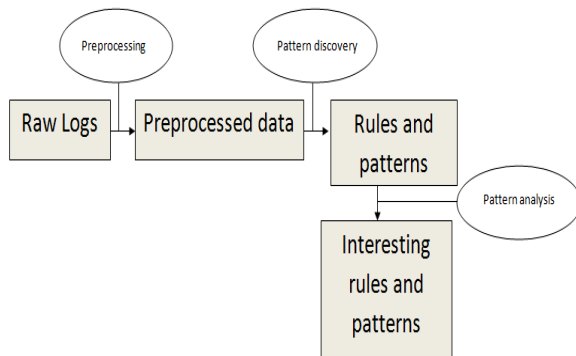

Fig2: Web Usage Mining Process

Preprocessing of text, image, scripts and multimedia files are carrying out in content preprocessing. Structure preprocessing involves the processing of hyperlinks between the page views.

4.2 Pattern Discovery

Pattern discovery is that set of methods, algorithms, and techniques used to extract patterns from web log file. Several techniques are used for pattern discovery such as statistical analysis, clustering, classification, and sequential pattern mining. After patterns are discovered they need to be analyzed in order to determine interesting and important patterns, besides the removal of redundant patterns [8]. Pattern analysis has several different forms such as knowledge query mechanism, visualization techniques, and loading usage data into a data cube in order to perform Online Analytical Processing OLAP operations. A web server log file records user's transactions in the web. Usually, the web log file contains information about the user IP address, the requested page, time of request, the volume of the requested page, its referrer, and other useful information.

4.3. Pattern Analysis

Pattern analysis is the last part of Web Usage Mining. This phase will filter out all unimportant patterns from the set found in the pattern discovery. Knowledge query mechanism, such as SQL, is the most common form of pattern analysis method. These use content and structure information also for filtering out patterns containing pages of certain usage types, content types or pages that match a certain hyperlink structure.

Data mining techniques mostly used in web usage mining are statistical analysis techniques, clustering, classification, association rule mining, and sequential pattern mining. Statistical analysis is the process of applying statistical techniques on web log file to describe sessions, and user navigation such as viewing the time and length of a navigational path. Statistical prediction can also be used to predict when some page or document would be accessed from now. It makes use of the N-grammar model which assumes that when a user is browsing a given page, the last N pages browsed affect the probability of the next page to be visited.

Clustering is the process of partitioning a given population of events or items into sets of similar elements. In web usage mining there are two main interesting clusters to be discovered: usage clusters, and pages clusters. An approach is to cluster web pages to have high quality clusters of web pages and use that clusters to produce index pages, where index pages are web pages that have direct links to pages that may be of interest of some group of website navigators.

Classification is dividing an existing set of events or transactions into another predefined sets or classes based on some characteristics [1]. In web usage mining, classification is used to group users into predefined groups with respect to their navigation patterns in order to develop profiles of users belonging to a particular class or category.

Association rule mining is the discovery of attribute values that occur frequently together in a given set of data. Association rules mining techniques are used in web usage mining to find pages that are often viewed together, or to show which pages tend to be visited within the same user session. A re-ranking method with the help of website taxonomy is to mine for generalized association rules and abstract access patterns of different levels to improve the performance of site search. Another approach for predicting web log accesses is based on association rule mining.

Association rule mining facilitates the identification of related pages or navigation patterns which can be used in web personalization.

5. Characteristics of data mining in terms of E-commerce:

- **Availability and Partition tolerance, with tunable consistency**: We should be able to trade off on consistency to accommodate highly

available and partition tolerant engagement services.

- **Linear Scalability**: In the face of massive amounts of content being created by users the data mining system should be able to scale without degrading performance.
- **Passable Code Base**: The data mining code base should compromise both in size and complexity,
- **Aggregators are a first class concern**: aggregator modification is potentially the most heavy write load element of an engagement service. So the Data mining should support high performing aggregator over a distributed infrastructure.
- **Sorting is a first class concern**: Most user generated content is going to be in a sorted form, ranging from most recent comments (like news feed) to sort by helpfulness. Sorting large amounts of data should be handled by data mining in a highly efficient manner.
- **Multiple pivot points for data elements**: each complex data entity should be accessible through its attributes through a reverse index or filtering.
- **Search integration**: text search should either be native or easily pluggable into the data mining engine.
- **Selectable individual attributes**: Attributes of a data entity should be individually selectable and updatable.
- **Schema Less**: The data model should be flexible and should not impose any constraints on what and how much data is stored. A schema less data store like columnar or key-value data stores provide great data model flexibility.
- **Native support for ephemeral data**: E-commerce services are going to generate lots of data of an ephemeral nature, i.e. data that is important only for a short period of time. Ephemeral data should not clog up the system.
- **Replication should be first class concern**: replication, replication awareness should be deeply integrated into the design and implementation of the data warehouse.

6. Conclusion

In this paper, we have presented how web mining (in a broad sense, data mining applied to E-commerce) is applicable to improve the services provided by e-commerce based enterprises. Specifically, we first discussed the interesting pattern analysis where the usage of association rules has been discussed where some additional constraints are highlighted. We then reviewed latest recommender systems based on

location of the customer. Later we have presented the web usage mining process. At the last we have presented the latest challenges faced by data mining in terms of e-commerce.

References:

[1] Rajesh Natarajan and B Shekar ," Interestingness of association rules in data mining: Issues relevant to e-commerce", *Sadhana* Vol. 30, Parts 2 & 3, April/June 2005, pp. 291–309.

[2] Hamid Rastegari, Mohd Noor Md. Sap, Jurnal Teknologi Maklumat "Data Mining and E-Commerce: Methods, Applications, and Challenges in December 2008.

[3] Istrate Mihai "WEB MINING IN E-COMMERCE".

[4] Aditi Todi, Dr. Rajashree Shettar "Classification of E-Commerce Data Using Data Mining", international journal of engineering science & advanced technology Volume-2, Issue-3.

[5] Rupesh Sanchati, P.C. Patidar, Gaurav Kulkarni "Path Breaking Case Studies in E-commerce Using Data Mining" International Journal of Computer Technology and Electronics Engineering Volume 1, Issue1.

[6] Kohavi, R. & Provost, F. (2001). Applications of data mining to E-commerce. In *Data Mining and Knowledge Discovery*, 5(1/2). Kluwer Academic.

[7] Kyoung-jae Kim, Hyunchul Ahn, and Sangwon Jeong "Context-aware Recommender Systems using Data Mining Techniques", World Academy of Science, Engineering and Technology 2010.

[8] Anupama Prasanth "Web Usage Mining – Its Application in E-Services", International Journal of Emerging Technology and Advanced Engineering Volume 3, Issue 2, February 2013.

[9]L.Niu, X.W.Yan, C.Q.Zhang and S.C.Zhang,"product hierarchy-based customer profiles for e-commerce recommendation", Proceedings of International conference on Machine Learning and Cybernetics, vol 2,pp1075-1080.