International Journal of Engineering & Technology, 7 (1.1) (2018) 286-289

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Personalized web search on e-commerce using ontology based association mining

B. Sekhar Babu [1] *, P. Lakshmi Prasanna [2], P. Vidyullatha [3]

[1] *B. Sekhar Babu, Department of Computer Science and Engineering, KLUniversity*
[2] *P. Lakshmi Prasanna, Department of Computer Science and Engineering, KLUniversity*
[3] *P. Vidyullatha, Department of Computer Science and Engineering, KLUniversity*
**Corresponding author E-mail: sekhar99@gmail.com*

Abstract

In current days, World Wide Web has grown into a familiar medium to investigate the new information, Business trends, trading strategies so on. Several organizations and companies are also contracting the web in order to present their products or services across the world. E-commerce is a kind of business or saleable transaction that comprises the transfer of statistics across the web or internet. In this situation huge amount of data is obtained and dumped into the web services. This data overhead tends to arise difficulties in determining the accurate and valuable information, hence the web data mining is used as a tool to determine and mine the knowledge from the web. Web data mining technology can be applied by the E-commerce organizations to offer personalized E-commerce solutions and better meet the desires of customers. By using data mining algorithm such as ontology based association rule mining using apriori algorithms extracts the various useful information from the large data sets .We are implementing the above data mining technique in JAVA and data sets are dynamically generated while transaction is processing and extracting various patterns.

Keywords: *Association; Datasets; Ecommerce; Ontology.*

1. Introduction

Current growth of web technology, the E-Commerce has changed into one of the most important Networks in the recent style of purchasing. By observing and recognizing the performance and desires of the E- customers, the e-commerce organizations can acquire an awareness to create and make relationship planning with the buying across the internet and offer personalization solution. Personalized marketing has become an important business issue for e-commerce traders in past few years. On the e-commerce internet site, there exists lots of on-line transactions, generating a huge number of documents and records of the cataloguing form every day. Ecommerce organizations are challenged with a richness of data, lack of information of awkwardness.

Recommender systems have changed into a primary web-application, due to simplicity of collecting the personal information on web. The most important difficulty in additional expansion of e-commerce has been shortage of appropriate personalized information about the users. Personalization in e-commerce web sites defines the traders in contributing the exclusive products and appropriate services that depends on every purchaser's specific information like age, status, profession, his/her purchasing behavior and preference.

Through personalized information service system, the e-commerce organizations can provide customers with information content and system tasks fulfilling their individual desires. The basic information and access designs related to buyer's can be mined using personalized service in the creation of pricing strategies based on market strategies.

In this situation pretending the personalization of trade becomes significant due to habits, trends and demand could be predicted to analyze the appropriate customer's groups or the individual's forthcoming buyer behavior. This leads to the decision of specific marketing, cost savings, improving efficiency, so as to get more benefit for traders. Web data mining can be used in e-commerce applications to find data that have high value.

The E-Commerce web sites are turning into very large and more complex. Therefore it is usually very hard for purchaser's to choose among huge number of goods. So how to advice the user's with satisfactory individual recommendations is repeatedly being measured as a motivation of the new marketing approach in enterprise. Personalized Recommendation system exploits statistic method and knowledge discovery technology to resolve the recommendation problems when communicating with the customers. It gives the probabilities and suggestion about the products as what a salesman does until the purchasing is satisfied. The same technology is being applied in personalized recommendation systems.

There are mainly two types of web mining which is stated as web content mining and web usage mining. Web content mining involves in frequently searching information resources in web pages and web usage mining is associated to determine user access patterns from web usage data. Web structure mining determines useful knowledge from hyperlinks which signifies the structure of the Web and its content. Web usage mining analyses user's access to the Web pages through mining the Web log files and associated data.

E-commerce transactions generate large amounts of data as by gathering , processing and managing the huge amount of data associated to customer behavior, accurate user's groups or individuals can create an efficient personalized model and reduces the load of Dynamic web-pages on E-commerce sites. Finding the maximum confidence and support [6].

Copyright © 2018 B. Sekhar Babu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data mining is defined as the mining of the unknown information from the available past information and making use of it for estimating the upcoming actions. Data mining in modern days has been executed in several industries such as retail, manufacturing, health care, banks and tele communication [5]. From last several years data mining has taken the very important roles in the sales of the merchants and also helped to expand the benefits by introducing statistical and computational techniques [5]. Data mining helps in defining the patterns by using those statistical methods such that the retailer can distinguish about the buying percent of every item, such that the retailer can remove the unwanted products. Based on the patterns mined the sellers can have an structured arrangement of the items in the store such that the buyer will show attention in buying the products for the very next time. Hence the hit count rate of the website also grows the additional profit for the enterprise by encouraging some announcements on the website [5].

Another technique in data mining is classification of the existing data. The classification of data is done by using some fixed attributes such that we can decrease the data to be processed for each and every transaction. This profits in recognizing the interest of the customer based on his attributes taken from his past transactions from the website [1]. Mainly the classification is used for estimating in mining of the given sales data. By considering the labeled data and classifying the data according to appropriate category will generate some patterns about that data taken such that the up-coming events are recognized and processed accordingly. Association rule mining is used find the frequently sold items and the relationship between those items, the outlier items that have to be excluded from the stores. The association means the direct relation between any two attributes and their dependencies [6]. This is a computational method where we find the parameters such as support and confidence. These are the two which decides the type of relationship between the items and the customers. Not only a high confidence and support is helpful but also a low support and confidence to determine an unusual behavior of transactions can predict the fraud from the customers. Apriori algorithm is used for mining the item sets that occur frequently by some Boolean expressions on the data set given. Apriori algorithm is an iterative search algorithm finds the strong relations by finding the maximum confidence and support [6].

Java is platform is used for the developing web pages dynamically hence whole data deposited can be stored using MY SQL and HTML pages are designed in order to provide more efficient interactive web pages .Therefore statics and analysis of the data is done by applying the data mining technique. Many data miners use JAVA platform for their knowledge mining techniques for faster results [4]. JAVA is also a language available in several forms and also to design jsp pages can combine different formats of statistical methods into a single program. In java the algorithms are written in in servlet programs such that the users can easily the code and extract the required knowledge. JAVA consists of some in built functions and libraries such that the default cases are avoided from the user to write large programs. Java can be applied on variables and the information can be stored in the memory as object. JAVA p is an interpreter therefore every line of the written program is perfectly checked and processed. JAVA is a simple and easy language and can have a simple syntax that can be easily understood by the beginners [4]. The operations are done on the active memory of the computer. For getting the statically information and strong web mining we need to prefer some built in algorithms to do in the JAVA and it is known for its simplicity. The data analyzed as result is directly shown on the screen [8].

Every approach in this paper is done by using JAVA platform by loading the data dynamically cvs file into the JAVA. By applying ontology based association mining by using JAVA built in packages are used to obtain certain patterns a[4]. The statistical information that is taken from the result of the algorithms specifies the retail traders to perform the future process. The strength of JAVA is that we can program our algorithms easily that is understood by every user. It also produces the mathematical formulas required that are in built in R for pre-processing. The R not only produces statistical and computational results but also gives out the strong associations among the frequent item set [8].

2. Apriori algorithm

Apriori algorithm is used for the frequent item set mining by using the Boolean values. The subset of a frequent item set must be frequent is the main aim of the apriori algorithm. For example let us consider an item set consisting of list of items and their transaction id's [3]. Now the minimum support and confidence are specified to exclude unwanted items and improve the efficiency of producing the frequent item sets based on the support and confidence computational results [3]. This process is applied recursively until we find the association list with same support and confidence and organize all the items in the same place or give out the recommendation of remaining products on buying one product from the associated item list [3].

Step 1: Consider a data set with transaction ids.
Step 2: Specify the minimum support and confidence.
Step 3: Then find the frequent items by using this confidence and support. Exclude the items that doesn't support the minimum support and confidence.
Step 4: Next iterate the same process by taking the combinations of the selected items.
Step 5: Repeat this process until there is no set with association.
Step 6: Find the set that is associated and organize them in the same place.

3. Ontology based association mining

Ontology is a demonstration of official knowledge. It gives a clear and Consistent representation of language and techniques that help individuals to detect the difficulties and dealing with affairs, deliver public languages of areas and determine various levels of formal meanings of terms and relationships between terms. It is systematized by taxonomy and comprises the typical model of the native language of the ontology and can provide a public and reliable understanding of the area. It
overcome the semantic content of the Communication and mismatch problem. Ontology structure is classified into the following six levels:

1) Recognize the purpose and aim of the ontology application: Begin the area of study Ontology Can be established based on the Ontology of domain
2) Ontology Investigation: Determine the association among all kinds of Ontology and terms.
3) Ontology Representation: It can Choose with the suitable method of ontology allowing to the System necessity.
4) Ontology testing: It can test the quality, stability, reliability, and its be scalable.
5) Ontology Structure: It can check the ontology permitting to the above principles, to encounter the necessities, where it can store the files, or else it will switch to next process.
6) Ontology representation: By representing Ontology many of the languages are in exist.

Applications of Ontology --based Web Mining

1) Improve search to Web
2) It gives the better browse capability
3) Personalize with Web data access.

4. Structure of iwum

Iterative Web Unit Mining Method

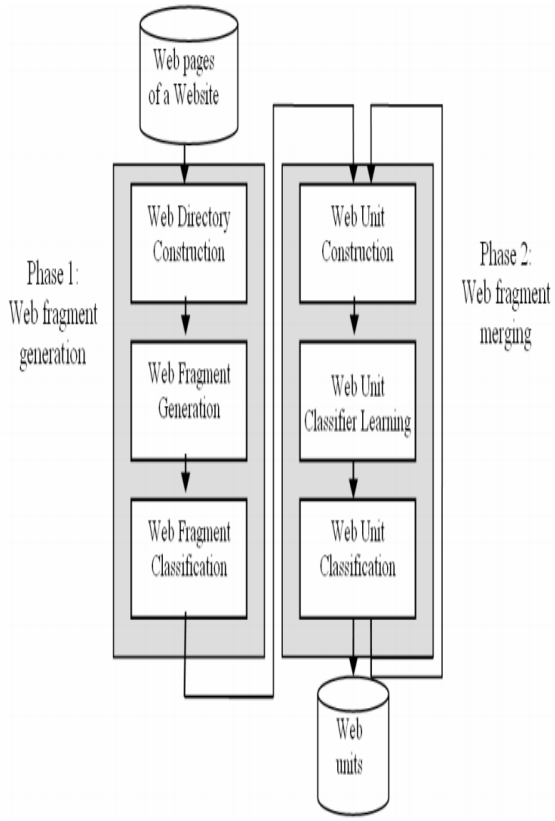

Fig. 1: Structure of IWUM.

5. Platforms used for implementing data mining techniques

There are various platforms where we can implement all the data mining techniques for analysis of big data are Hadoop, R, weka and python etc. Many organizations use Hadoop for better analysis of their data [8]. By using the Hadoop for any business we can take the input data, process the data, manage the data, analyzing the information and storing the required statistics for the fast organization growth. For doing the predictions from the given data we use weka for real time applications [8]. Weka uses a graphical user interface and the command line interface. Therefor it becomes easier for an analyst to process in weka. Python is a programming language which is easy to learn even by the learners or beginners, it is a flexible language that can be used for analysis and manipulation of data [8].

5.1. CVS file data

Successfully Created CSV file : sfokzqna.c
Contents of CSV file : sfokzqna.csv
lipstick,Nuts,Carrots,HugoDarkBlue
Onions,Milk-Blue,Grey
BeetRoot,lipstick,Onions,Salt,Bread,ROYAL
Carrots,Salt,ARAMIS
Nuts,Onions,ROYAL
Carrots,Salt,ROYAL
BeetRoot,Nuts,ARAMIS
Onions,Milk-Orange,Grey
Nuts,Onions,ROYAL
lipstick,Onions,Milk-Orange
Onions,Grey

RRAR RuleMining

5.2. Calculating the confidence for each transaction

All transcastions loaded into memory.

[Nuts, ROYAL], 18.181818181818183%
[Onions, lipstick], 18.181818181818183%
[Onions, Grey], 27.272727272727273%
[Salt], 27.272727272727273%
[Salt, ROYAL], 18.181818181818183%
[Onions], 63.636363636363636%
[ARAMIS], 18.181818181818183%
[Salt, Carrots], 18.181818181818183%
[Nuts], 36.363636363636373%
[Onions, Nuts], 18.181818181818183%
[Milk-Orange], 18.181818181818183%
[Grey], 27.272727272727273%
[Milk-Orange, Onions], 18.181818181818183%
[Carrots], 27.272727272727273%
[ROYAL], 36.363636363636373%
[Onions, ROYAL], 27.272727272727273%
[Onions, Nuts, ROYAL], 18.181818181818183%
[BeetRoot], 18.181818181818183%
[lipstick], 27.272727272727273%

==High-confidence association rules(min_conf=30.0%)
[ROYAL] => [Salt] (conf:50.0%)
[Carrots] => [Salt] (conf:66.66666666666667%)
[lipstick] => [Onions] (conf:66.66666666666667%)
[Grey] => [Onions] (conf:100.0%)
[Nuts] => [Onions] (conf:50.0%)
[Milk-Orange] => [Onions] (conf:100.0%)
[ROYAL] => [Onions] (conf:74.99999999999999%)
[Nuts, ROYAL] => [Onions] (conf:100.0%)

5.3. Ontology web rule mining RARR data set

All transcastions loaded into memory.

==High-confidence Associations
[Salt], 21.4285714285714 27%
[Onions], 50.0%
[lipstick], 21.4285714285714 27%
[BeetRoot], 14.2857142857142 85%
[Onions, Nuts], 14.2857142857142 85%
[Nuts], 28.571428571428 57%
[Salt, Carrots], 14.2857142857142 85%
[Carrots], 21.4285714285714 27%

==High-confidence association rules
Medium Rules: [Carrots] => [Salt] (conf:66.66666666666666%)
Medium Rules: [Salt] => [Carrots] (conf:66.66666666666666%)

6. Conclusion

Personalization aims to build the model based on user's information gathered during the Transactions carried out on the E-commerce sites. Personalization always plays a key role in success of E-commerce trade because personalization attracts the customers towards the E-Commerce sites and improves the buying patterns. Personalization means of meeting the customer's desires more effectively and efficiently. Hence we built the personalization model by using data mining techniques such as ontology based association rule mining in JAVA platform to predict the customers behavior, purchasing habits, based on various parameters such as age, status, occupation, gender, brand etc..,we can easily identify which age group persons buys what find of product and which brand is rating more in market ,sales analysis for various items and pricing limits for various customers based on their based on their occupation and also identifies which item sets are frequently purchased by the customers. Hence provides the rough scenario to predict the customer's needs and can provide the personalization when new user logs on the e-commerce sites. There by e-commerce sites can increase their sales on various products and can provide huge profits to the E-commerce organization

References

[1] B.Sabitha1, N.G.BhuvaneswariAmma2G.Annapoorani3, P.Balasubramanian4, Implementation of datamining techniques to perform market analysis vol2, issue 11.

[2] Esteban Alfaro, Mat´ıas G´amez, Noelia Garc´ıa, An R Package for Classification with Boosting and Bagging, vol54, issue2.

[3] Market basket analysis using fp growth and apriori algorithm: a case study of mumbai retail store prof. kavitha venkatachari.

[4] The art of R programming by norman matloff.

[5] Data science and big data analytics by Wiley.

[6] Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka A. M. Khattak, A. M. Khan, Sungyoung Lee* , and Young-Koo Lee.

[7] Role of data mining in retail sector bharati m. ramageri

[8] An overview of free software tools for general datamining,

[9] Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions Pramod Prasad, Research Scholar

[10] Bigdata analytics in retail and consumer services.

[11] Zhang Haiyang "The research of Web mining in ecommerce" International Conference on Management and Service Science (MASS), 2011.

[12] Yanguang Shen, Lili Xing,Yiting Peng "Study and application of Web-based data mining in e-business"Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, SNPD 2007

[13] Mark Xu, John Walton, (2005) "Gaining customer knowledge through analytical CRM", Industrial Management & Data Systems" https://doi.org/10.1108/02635570510616139.

[14] Shen Zihao, Wang Hu "Research on e-commerce application based on Web mining" International Conference on Intelligent Computing and Cognitive Informatics (ICICCI), 2010. https://doi.org/10.1109/ICICCI.2010.89.

[15] Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions Pramod Prasad, Research Scholar.

[16] Francisco, G. S., Rafael. An integrated approach for developing e-commerce applications. Expert Systems with Applications, 2005, 28:223–235. https://doi.org/10.1016/j.eswa.2004.10.004.

[17] Schafer, J. B., Konstanz, J. Recommender systems in Ecommerce. Proceedings of 1st ACM conference on electronic commerce, Denver, Colorado, USA, 1999, 158-166. https://doi.org/10.1145/336992.337035.

[18] S. Hung, C. Lin and J. Hong, "Web mining for event-based common sense knowledge using lexicons-syntactic pattern matching and semantic role labelling," Expert Systems with Applications, vol.37, pp. 341-347, 2010. https://doi.org/10.1016/j.eswa.2009.05.060.

[19] P. Chou, P. Li, K. Chen and M. Wu, "Integrating web mining and neural network for personalized e-commerce automatic service," Expert Systems with Applications, vol.37, pp. 2898-2910, 2010. https://doi.org/10.1016/j.eswa.2009.09.047.

[20] ZHOU Hoofing, ZHU Jacqui, ZHU Yang Yonge, and SHI Bailee, "Armine: A Data Mining Tool Based on Association Rules,"Journal of Computer Science and Technology", vol. 17, 594-602, Jan.2002. https://doi.org/10.1007/BF02948827.

[21] Jindong Liu. WEB data mining research applications in Ecommerce [J]. Network security technology and Application, 2013 (9): 25-26.

[22] Xianchun Zou, Zhonghu Xie, Yanbu Zhou. El-commerce and Web data mining [J]. Computer application, 2001, 21(5): 21-23.

[23] Li Kan, Hao Pan. Application of Web data mining technology in E-commerce [J]. Computer knowledge and technology, 2010, 4:816-81.

[24] Ru Nie, Senping Tian. Web data mining and its application in Ecommerce [J]. Beijing: Computing technology and automation, 2006, 12: 237-239.

[25] basher, B. et al, Automatic Personalization Based on Web Usage Mining, Comm. of the ACM,Vol43, No8, Aug, 2000.

[26] 26. Manu rice D, Mulvane et al, Personalization on the Net using Web Mining, Comm. of the ACM, Vol43, No8, Aug, 2000.

[27] Udi Manber et al, Experience with Personalization on Yahoo, Comm.

[28] Chang, ByeongCheo, "A Study on Multi-Criteria Individualized Commodity Recommendation in-Shopping Mall", the Graduate School of Sungkyunkwan University, 2002.

[29] Liu kai Nan, "Study on e-commerce data mining technology based on the interest," Market Modernization, January 2007.

[30] Ji Fang-fang, "Analysis of the development of e-commerce in China," Jiangsu Commercial Forum, March 2008.