



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

A Survey on Web Mining In E-Commerce: Pattern Discovery, Issues and Applications

AkashKalmegh¹, Atul Raut², Shubham Sonule³, Shubham Shivade⁴, Ass. Prof. B. S. Satpute⁵

Student, Department of Computer Engineering, Pad. Dr.D.Y.Patil Institute of Engineering and Technology,

Pimpri, Pune, India^{1,2,3,4}

Department of Computer Engineering, Pad. Dr.D.Y.Patil Institute of Engineering and Technology, Pimpri, Pune,

India⁵

ABSTRACT: In last two decade the speedy growth of e-commerce and the huge amounts of data collected through operational transactions, data mining techniques are becoming more useful to generate and understand unknown customer patterns. In the past, data mining has been used to find out which products are related in terms of having maximum buy and also ascertain which customers worthy credit facilities. There has not been much work done in the use of data mining to ensure customer allegiance in the e-commerce business and also have strategies of increasing retail companies to use e-commerce as a profitable mode of doing business. The aim to study the customer's behavior through data mining techniques used in deriving association rules from an e-commerce database so as to ensure customer loyalty and also assist in having strategies of luring businesses to use e-commerce for conducting highly profitable business. From our results the association rules reveal that if a product stays online for a long time (more than 550 days), it is 78% highly likely it will not be bought. The associations rules also indicate that the number of products bought are linked to the number of times customers view the products online and the selling price of the product and compare a product from other ecommerce sites.

KEYWORDS: E-commerce, Association Rule, Patterns, Recommendation, knowledge discovery, Ranking, Comparison.

I. INTRODUCTION

Mining knowledge from big data and large databases has been recognized by many researchers as a key research topic in database systems and machine learning and by many industrial companies as an important area with an opportunity of stature revenues. Mining high utility item sets from a transactional database refers to the discovery of different item sets with high utility like profits and sales. Businesses that deal with electronic commerce (e-commerce) use web data mining to obtain useful information about their customers, products and competitors. The extracted information can help these businesses to decide their advertising strategies, promote targeted marketing and reduce operating costs. The knowledge about the business environment which includes understanding customers, market trends and customer feedback also helps software developers improve their website designs whilst optimizing the website structure and easy website navigation for customers. Techniques such as association rule mining permit the analysis of shopping cart data to improve the presentation or location of products. Data mining (DM) has attracted a great deal of attention in the information industry, the healthcare sector and in society as a whole in recent years, due to the wide availability of large amounts of data and the imminent need for turning such data into reliable information and knowledge. DM is applied to generate interesting patterns and knowledge from large databases. The knowledge and intelligence gained can be used for a wide range of applications which may include market analysis, customer retention, stock and production control, healthcare systems and science exploration. DM and knowledge discovery are commonly seen as intelligent tools that help to accumulate and process data and make use of it. DM bridges many technical areas, including databases, statistics, neural networks, machine learning, and human-computer interaction. The set of DM processes used to extract and verify patterns in data is the core of the knowledge discovery process.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

II. LITERATURE REVIEW

1. Data mining on social security and social welfare data (2014).

From this paper we Refer-

Data mining has attracted a great deal of concentrating on the information industry and in society as a whole in past years, due to the wide availability of large amounts of data and the immediate need for turning much data into useful information and knowledge. The government faces new and intense pressure to collect and use personal data. The need for a deep understanding of public and public-government influence through advanced data analytics has been increasingly recognized by the community at large. Mining Social security/welfare data is challenging. The challenges arise from business, data, and the mining of the data. Social Security Data Mining (SSDM) attempt to discover interesting patterns and exceptions in social security and social welfare data. The SSDM framework including business an research issues, social security/welfare services and data, and the different methods for SSDM framework.

2. Efficient Algorithms for Mining High Utility Item sets from Transactional Databases (2013).

From this paper we Refer-

Mining high utility item sets from a transactional database refers to the discovery of item sets with high utility like profits. Although a number of useful algorithms have been proposed in recent years, they incur the problem of producing a large number of candidate item sets for high utility item sets. Such a huge number of candidate item sets degrades the mining performance in terms of it require more execution time and large space requirement. The situation may become worse when the database contains lots of long transactions or long high utility item sets. In this we propose two algorithms, namely utility pattern growth (UP-Growth) and UP-Growth+, for mining high utility item sets with a set of effective strategies for trimming candidate item sets. The information of high utility item sets is maintained in a tree-based data structure name utility pattern tree(UP-Tree) such that candidate item sets can be generated efficiently with only two scans of database. The performance of UP-Growth and UP-Growth+is compared with the state-of-the-art algorithms on many types of both real and synthetic data sets. Experimental results show that the proposed algorithms, especially UP-Growth+, not only reduce the number of candidates effectively but also outperform other algorithms substantially in terms of runtime, especially when databases contain lots of long transactions.

3. Data Mining with Big Data

From this paper we Refer-

Big Data concerns with huge amount of data, complex, growing data sets with multiple, autonomous sources. With the rapid development of networking, data storage, and the data collection capacity, Big Data is now fastly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This article presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data driven model involves demand driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. We analyze the challenging issues in the data driven model and also in the Big Data revolution.

4. Predicting School Failure Using Data Mining (2013).

From this paper we Refer-

This system proposes to apply data mining techniques to tell in advance school failure. We have used real data about 670 middle-school students from Zacatecas, México. Several operations have been carried out in an attempt to improve accuracy in the prediction of final student performance and, specifically, of which students might fail. In the first experiment the best 15 attributes has been checked. Then two different techniques have be seen applied in order to resolve the problem of classifying unbalanced data by rebalancing data and using cost sensitive classification. The outcomes of each one of these approaches using the 10 classification algorithms and 10 fold-cross validations are shown and compared in order to select the best approach to our problem.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

5. Data Mining in Electronic Commerce

From this paper we Refer-

Modern business is moving hurriedly toward e-commerce. If the transition is done properly, it enables better management, new services, lower transaction costs and better customer relations. Success depends on skilled information technologists, among who are statistician's . In this focuses on some of the contributions that statisticians are making to help change the business world, especially through the development and application of data mining in systematic way. This is a very big area, and the subject we cover are chosen to avoid overlap with other system in this special issue, as well as to respect the limitations of our expertise. Inevitably, electronic commerce has raised and is raising fresh research problems in a very wide range of statistical areas, and we try to emphasize those challenges.

III. SYSTEM ARCHITECTURE

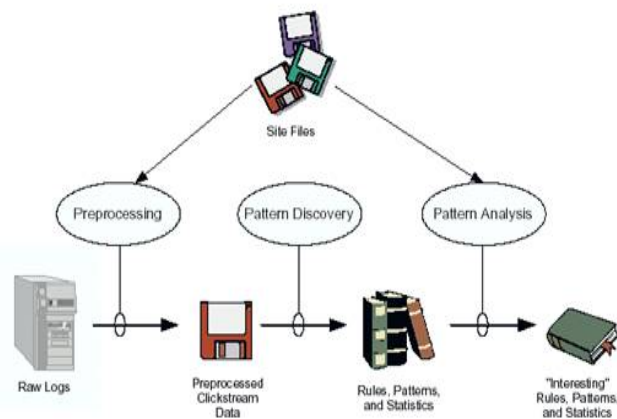


Fig No 1: System Architecture

IV. WORKING/METHODOLOGY

The useful information can help these businesses to decide their advertising strategies, promote targeted marketing and reduce operating costs. The knowledge about the business environment which includes understanding customers, market trends and customer feedback also helps software developers improve their website designs whilst optimizing the website structure and easy website navigation for customers. Techniques such as association rule mining permit the analysis of shopping cart data to improve the presentation or location of product. There is a large body of work on image content analysis, for classification and interpretation and photo ranking , also in the context of online photo sharing sites, such as Flickr . Of these works, Zerr's work is probably the closest to ours. Zerr explores privacy aware image classification using a mixed set of features, both content and meta-data. Wemap the user collective policy specification to an auction based on the Clarke-Tax mechanism which selects the privacy policy that will maximize the social utility by encouraging truthfulness among the co-owners. We are going to study the customer's behavior through data mining techniques used in deriving association rules from an e-commerce database so as to ensure customer loyalty and also assist in having strategies of luring businesses to use e-commerce for conducting highly profitable business. From our results the association rules reveal that if a product stays online for a long time (more than 550 days), it is 78% highly likely it will not be bought. The associations rules also indicate that the number of products bought are linked to the number of times customers view the products online and the selling price of the product.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

1. User Personal Interest:

Besides the trust values between friends in the same category, user interest is another significant factor to affect users' decision-making process, which has been proved by psychology and sociology studies. Moreover, demonstrated the effect of Context MF model with consideration of both individual preference and inter personal influence. However, there are two main differences of the user interest factor in our model to individual preference in Context MF:-

1) *The independence of user interest. It means we can recommend items based on user interest at a certain extent. In other words, we utilize user's connection with the items to train the latent feature vectors, especially for the experienced users.*

2) *Interest circle inference .Just like Circle Con model, we divide the tested social network into several sub-networks, and each of them corresponds to a signal category of items. Considering the cold start users who has a few rating records, we use friends' interest in the same category to link user latent feature vector.*

2. User Interest Description:

According to the natural item category tags of rating datasets, we can get category distribution of the item, which can be seen as the naïve topic distribution of the For example, each item has the tags of its category in Yelp. Just like the item The Dakota Bar of New York belongs to the category Night Life, and then Night Life is one of the category tags of the item. From user's historical rating data in category c, we summarize all the rated items to measure user interest.

3. Personal Interest:

Due to the individuality, especially users with many rating records, users usually choose items all by themselves with little influence by their friends. However, many previous works took the circles of friends in social networks to solve the cold start problem. It did work for the cold start users with a few records, but ignored the individuality for experienced users. In other words, the relevance of user and item latent feature vector depends on the relevance of user interest D_u and item topic D_i to a certain extent. More formally, we denote the relevance of user personal interest to the topic of item in our recommendation model.

4. Interest Circle Inference:

Similar to the trust circle inference in Circle Con model, we propose the interest circle inference. The basic idea is that user latent feature vector should be similar to his/her friends' latent feature vector based on the similarity of their interest.

V. CONCLUSION

This system has attempted to provide an up-to-date survey of the rapidly growing area of Web usage mining, which is the demand of current technology. In this a general overview of Web usage mining is presented in introduction section. Web usage mining is used in many areas such as e-Business, e-CRM, e-Services, e-Education, e-Newspapers, e-Government, Digital Libraries, advertising, marketing, bioinformatics and so on. The major classes of recommendation services are based on the discovery of navigational patterns of users. The main techniques for pattern discovery are sequential patterns, association rules, Classification, Clustering, and path analysis. Web usage mining's basic components, taxonomy of web mining, architecture of web usage mining, individual components in web usage mining and detailed research.

ACKNOWLEDGEMENT

We would like to acknowledge and extend our heartfelt gratitude to our guide Ass.Prof.B.S.Satpute,DYPIET for his expert guidance and encouragement.

REFERENCES

- [1] M. S. Chen, J. Han and P. S. Yu, "Data mining: An overview from database perspective," IEEE Trans. on Knowledge and Data Engineering, vol. 8, no 6, pp.866-883, December 1996.
- [2] V. S. Tseng, B. E. Shie, C. W. Wu and P. S. Yu, "Efficient algorithm for mining high utility itemsets from transactional databases," IEE Trans. on Knowledge and Data Engineering, vol. 25, no. 8, pp.177 1786, August 2013.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

- [3] D. L. Banks and Y. H. Said, "Data mining in electronic commerce Statistical Science, vol. 21, no. 2, pp.234-246, May 2006.
- [4] L.-J. Kao, Y.-P. Huang and F.E. Sandnes, "Associating unobserved frequent itemsets with infrequent items to identify abnormal transactions," Applied Intelligence, vol. ??, no. ??, pp.??-??, November 2014.
- [5] M. S. Chen, J. Han and P. S. Yu, "Data mining: An overview from database perspective," IEEE Trans. on Knowledge and Data Engineering, vol. 8, no 6, pp.866-883, December 1996.
- [6] V. S. Tseng, B. E. Shie, C. W. Wu and P. S. Yu, "Efficient algorithm for mining high utility itemsets from transactional databases," IEEE Trans. on Knowledge and Data Engineering, vol. 25, no. 8, pp.177-1786, August 2013.
- [7] D. L. Banks and Y. H. Said, "Data mining in electronic commerce Statistical Science, vol. 21, no. 2, pp.234-246, May 2006.
- [8] L.-J. Kao, Y.-P. Huang and F.E. Sandnes, "Associating unobserved frequent itemsets with infrequent items to identify abnormal transactions," Applied Intelligence, vol. ??, no. ??, pp.??-??, November 2014.