RESEARCH HUB – International Multidisciplinary Research Journal (RHIMRJ)
Research Paper
Available online at: www.rhimrj.com

Review Paper introducing E-commerce, Data Mining and Pattern Recognition to recommend the product

Bhrantav Vora
Assit. Prof. Dept. of MCA
LDRP Institute of Technology and Research
Kh – 5, Sector 15, Near Gandhinagar ITI, Gandhinagar

Dr. N J Patel
Professor, Dept. of MCA
Acharya Motibhai Patel Institute of Computer Studies
Ganpat University, Kherva

Dr. N N Jani
Dean, Dept. of Computer Science
KSV University,
Gandhinagar

Girish Dave
Assit. Prof. Dept. of BCA
B P College of Computer Studies,
Sector 23, Gandhinagar

Abstract: ***E-commerce has become very prevalent for customers to buy goods and for vendors to sell and advertise the product. The study examined the criteria to increase the sales and recommend the appropriate product to the customer. The aim of this study is to introduce the e-commerce platform and the role of data mining technique and frequent pattern mining in e-commerce. A few feedback techniques including implicit, explicit and hybrid are reviewed from few review papers and their outputs are discussed. Data collection from different sources and then implementing data mining on that avail data is discussed. The recommendation system is highly depended up on the data collection which involves information collection phase, learning phase and prediction phase. This paper also discuss frequent pattern mining as a separate technique of data mining, frequent pattern mining with traditional framework, data stream, big data, uncertain data , graph and structured data are discussed.***

Keywords: Pattern Recognition, Data Mining, Frequent Pattern Mining.

I. INTRODUCTION

The term E-commerce or e-business refers to the business transactions among companies, (B2B) or the business transaction between companies and their customers (B2C) which will be completely or partially conducted over the internet or similar computer networks (Turban et al. 2015). The business completely depends on the information advantages. Gaining of more customers, relishing and retaining existing customers, and the appropriate prediction of the buyer behaviour will further enhance the availability of products and services and hence, the profit (Witten et al. 2016).

With the rapid development of internet and the information technology, e-commerce which is based on the virtual economy has been widely attracted and it has gradually developed into the backbone of the emerging industry. The network infrastructure was widely utilized, and the e-commerce development speed has been improved (Qin et al. 2014). While the e-commerce market scale expanded, more product manufacturers wanted to join the e-commerce market. An aggressive competition between the businesses develops as a result.

Enhancement of the profit rate in the intense market competition was a key issue. It has been observed that providing suitable product to the customers are of great importance (Persson et al. 2016). Hence, the e-commerce platform should quickly recommend the good products to the customers in terms of quality and the customer requirements. Thus,the e-commerce recommendation system is developed to provide the service of personalized recommendation based on different products to several customers (Li et al. 2015).

The e-commerce recommendation system focuses on user preferences, and allows users to choose the products they require. An excellent e-commerce product recommendation system successfully reduces the time cost for user to select the products according to their wish. In a recommendation system, a user may only rate a small number of items and therefore the rating matrix will be sparse. The user must estimate the missing rates to attain the knowledge of the proper recommendation results (Lu et al. 2015).

II. RESEARCH STUDY AREA

1.1 Recommendation system

The e-commerce sites make use of the recommender systems to suggest the best products to their customers and also to provide relevant information about the particular product to the consumers (Yoon et al. 2013). A particular product on the internet could be recommended based on the top sellers on a site, by considering the demographics of the consumer, or by conducting an

  ISSN: 2349-7637 (Online)

analysis of the past buying behaviour of the customers which acts a prediction for future buying behaviour (Hsu et al. 2015). There are several forms of recommendations. They include suggesting products to the customer, providing personalized product information, short and snappy community opinion, and also proving community reviews. The recommender systems includes the processes that are conducted largely by hand, like manually creating cross-sell lists, and also the actions that were performed largely by the computer (Heinrich et al. 2015). The recommender system advances the e-commerce in three ways,

1.1.1 Conversion of browsers to buyers: Mostly, the visitors of website often look at the website without purchasing anything. Here, the recommender systems could help consumers to find out the products they wish to purchase.

1.1.2 Increase in cross-sell: The recommender system improves the cross cell by suggesting additional products of customers choice. If the given recommendations are well and good, the average order size increases. The major example is suggesting of more products in the checkout process of a shopping cart. The recommendation system analyses a database of the consumer preferences to overcome the challenges of segment based mass marketing by presenting each customer with a personal set of recommendations (Palanivel et al. 2014). The recommendation systems are the technologies that could help the businesses to decide whom to make an offer with. Such kind of systems allows the search engines and the advertising companies to suggest advertisements or offers to display based on the consumer behavior. Online recommendations were preferred as they give respond quickly to the customer's preferences.

1.1.3 Cross selling: The Product-Associated Recommendations Suggestive selling is effective only when the seller identifies the current interest of the buyer. Retailers arrange the products to enhance the cross-selling by placing the complementary items in a close contiguity. Online retailers can directly suggest products that are related to the customer interest. Several recommender applications use the perspective of a current product or numerous current products in order to recommend other products by using a variety of recommender methods. This is due to the variety of inputs that could be used to generate such recommendations which includes anonymous purchase history, customer purchase history, product attributes, and expert opinions (Witten et al. 2016).

1.1.4 Building Long Term Relationships: The main aim of the retail businesses is to develop a long term relationship with the customers who lead to higher lifetime values. Personalization based on a customer's history of preferences, and purchases is the most difficult type of personalization to implement. Personalization is already common in web advertising and is widely used in e-commerce. Deep personalization builds a customer relationship over time, leveraging the history developed to provide better recommendations. This deep personalization uses user-user interaction, attribute based systems along with a learning module to identify user interests. The rise in popularity of the review aggregating websites has led to an influx of the data on customer's preferences (Wang et al. 2013). The huge repositories of the user written reviews create opportunities for new type of recommendation system that could control the content embedded in the text. The customer preferences are widely ingrained in the review texts which have an amble amount of characteristics that could be exploited by a neural network structure. The optimization techniques are adopted to find the objectives of the pattern mining technique.

The rise in popularity of the review aggregating websites has led to an influx of the data on customer's preferences. The huge repositories of the user written reviews create opportunities for new type of recommendation system that could control the content embedded in the text. The customer preferences are widely ingrained in the review texts which have an amble amount of characteristics that could be exploited by a neural network structure. The optimization techniques are adopted to find the objectives of the pattern mining technique.

The recommendation systems are beneficial to both service providers and users. These systems minimize the transaction cost of finding and selecting the items in online shopping environment. The recommendation system also have proved to improve the decision making process and its quality. In e-commerce settings, the recommendation systems enhance the revenues, for the fact that they are efficient means of selling more products (Aznoli et al. 2017). In libraries, these recommendation systems support the users by allowing users to move beyond the catalog searches. Hence, the necessity to use efficient and accurate recommendation techniques within a system which provides relevant and dependable recommendations for the customers could not be over emphasized (Ricci et al. 2015).

There are several phases in recommendation systems. They are given as follows,

- Information collection phase
- Explicit feedback
- Implicit feedback
- Hybrid feedback
- Learning phase
- Prediction phase

1.1.5 Information collection phase

In information collection phase, the relevant information of users is collected to generate a user profile for the prediction tasks which includes user attribute, and behavior of the sources that the user accesses.

1.1.6 Explicit feedback

This explicit feedback system normally stimulates the user through the system interface to provide the ratings for items to create and advance the model. The accuracy of the recommendation depends on the quality of the ratings provided by the user. The explicit feedback requires more effort from the user and also it provides transparency into the recommendation process which results in a slightly higher perceived recommendation quality (Tejeda et al. 2014).

1.1.7 Implicit feedback

The implicit feedback system automatically infers the user's choices by monitoring different actions of users such as history of purchases, navigation history, and time spent on web pages, and content of email. The implicit feedback minimizes the burden of customer by inferring their user preferences.

1.1.8 Hybrid feedback

The strengths of both implicit and explicit feedback could be combined in a hybrid system to reduce their weakness and attain a good performing system. This is achieved by using an implicit data as a check on the explicit rating.

1.1.9 Learning phase

The learning algorithm is applied to filter and exploit the user features from the feedback collected in information collection phase.

1.1.10 Prediction phase

The prediction phase recommends what kind of items the user may prefer. This will be done either directly based on the dataset collected in the information collection phase.

1.2 Data Collection

The data collection is the major step for execution of the data mining algorithms. The data collection technique includes the data reduction and selection techniques in order to improve the efficiency of data mining algorithms. Once the relevant data is collected, the data must be stored in a data warehouse (Amatriain et al. 2015). Data that are collected manually uses various kinds of data about individual customers. The data collection is classified into two types namely, factual data and transactional data. The factual data includes demographic information such as name, gender, date of birth, address, income, and the social security number. The major issue in developing recommendation application is constructing comprehensive user profiles based on the data collected. Some of the e-commerce sites include user choice like amazon.com (Abbas et al. 2015). The data available in the log files is frequently used to determine what profile could be dynamically processed in the background. We can use the data for pattern discovery techniques such as clustering, association rule mining, and sequential pattern discovery. The log data were collected automatically by web and the application servers represents the navigational behavior of the one who visits.

1.3 Data Mining

The data mining is the process of generating non-obvious necessary information for the decision makers form large databases. The term data mining was used to describe the collection of analysis techniques used to infer the rules from huge datasets (Banaee et al. 2013). One such well known example of the data mining in e-commerce is the development of association rules relationship between the items that indicates relationship between purchasing of one item and other item. Data mining has emerged as an active research area for extracting the implicit and potentially useful information from large set of databases.

1.3.1 Frequent Pattern Mining

The frequent item set pattern miningevolved to discover useful patterns in the customer's transaction databases. A customer's transaction database is a sequence of transactions, where each of the transaction is known as an item set. Most of the frequent mining algorithms have been designed with the traditional support confidence framework that generates interesting kind of patterns (Khare et al. 2013). These patterns uses different types of measures, model negative rules, or uses constraint based frameworks to determine relevant patterns.

1.3.1.1 Frequent Pattern Mining with the Traditional Support Framework

The traditional support framework is designed to determine the patterns for which the raw frequency is greater than the minimum threshold. Even though this is a simplest way of defining frequent patterns, this model has an algorithmic convenient property, which is known as level wise property. This level wise property of frequent pattern mining is highly crucial because it enables the design of a bottom up approach for exploring space frequent patterns. The efficiency of frequent pattern mining algorithm could be gained in several ways i) reducing the size of candidate search space using pruning methods ii) improving the efficiency of counting using database projection. iii) Using more effective data structures for further compressed database representation (Aggarwal et al. 2014). In frequent pattern mining, both the memory and computational speed could be improved by sensible choice of the data structures. In case the patterns to be mined are huge, then the number of subsets of the frequent patterns could be extremely large. Hence, a number of techniques need to be designed to mine long patterns. Various methods were used to explore long patterns, so that their subsets could be pruned efficiently (Gage et al. 2013).

1.3.1.2 Interesting Frequent Patterns

ISSN: 2349-7637 (Online)

One of the major challenges in frequent pattern mining is that the rules formed may not be interesting, when the quantifications such as support and confidence are used. This is because those quantifications do not normalize for the original frequency of the underlying items. An item which occurs rarely in the database would also occur in item sets with lowest frequency. Several methods have been proposed for finding the interesting frequent patterns which normalize for the underlying item frequencies.

1.3.1.3 Frequent Pattern Mining in Data Streams

The data stream has become very popular due to its advances in hardware and software technology that could collect and transmit data continuously. The major role of data mining algorithm is to execute the algorithms in single pass. This could be significantly challenging as the frequent and sequential pattern mining methods were generally designed as level wise methods. There are two types of variants of frequent pattern mining methods available for data streams i) frequent items, and ii) frequent item sets. In frequent items, the item sets needs to be determined from a data stream in a single pass (Chen et al. 2014). Such approach is generally required while the total number of distinct items is too large to be held in a main memory. In frequent item sets, it is not assumed that the numbers of distinct items are too large.

1.3.1.4 Frequent Pattern Mining with Big Data

In frequent pattern mining, the big data poses several challenges. Major problem arises when the data is large enough to be stored in a distributed way. Significant costs were acquired in scuffling around the data across distributed nodes. These costs were known as data transfer costs. When the datasets are too long, the algorithm needs to be designed to both disk access constraint and data transfer costs.

1.3.1.5 Frequent Patterns in Graphs and Structured Data

Numerous chemical and biological data, XML data, and the web browsing behaviors could be represented as structural graphs. In such cases, the frequent mining is essential for making inferences in the data. This is due to the reason that frequent structural patterns provide important insights about the graphs. The pattern mining becomes challenging when the graphs were too large. Another case is the streaming scenario where one has to determine the dense patterns in graph stream.

1.3.1.6 Frequent Pattern Mining with Uncertain Data

Probabilistic data has become increasingly common over last years, as the methods have been designed in a way to collect the data of low quality. The attribute values in such data were probabilistic, which implies that those values were represented as probabilistic distributions. Several algorithms have been proposed for the uncertain frequent pattern mining.

III. OBJECTIVES OF THE RESEARCH

- To explore the possible research to be carried out in recommendation system.
- To investigate the present scenario of e-commerce and data mining in context of frequent pattern mining.
- To identify the issues and challenges of recommendation system'.

IV. CONCLUSION

So, Recommendation system based on customer's feedback and reviews brings more business through e-commerce platforms. This system also produces very large amount of data which if mined perfectly, then it provides accurate guess of customer's requirements and choices for the product As well as habit of customers to buy some kind of products periodically. Use of pattern mining is useful to give product recommendation bases on customers behavior of buying pattern and searching pattern over the e-commerce platform and it gives opportunity to e-commerce platform to sell products frequently to the customer who has such kind of habit to purchase it periodically.

REFERENCES

1. Turban, E., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2015). *Electronic commerce: A managerial and social networks perspective*. Springer.
2. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
3. Qin, Z., Chang, Y., Li, S., & Li, F. (2014). *E-commerce strategy* of industries. In E-commerce strategy (pp. 151-206). Springer Berlin Heidelberg.
4. Persson, S., Fridolfsson, H., & Holst, A. (2016). Strategy within E-commerce: The formation process.

  ISSN: 2349-7637 (Online)

5. Li, S. S., & Karahanna, E. (2015). Online recommendation systems in a B2C E-commerce context: a review and future directions. *Journal of the Association for Information Systems*, 16(2), 72.
6. Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74, 12-32.
7. Yoon, V. Y., Hostler, R. E., Guo, Z., & Guimaraes, T. (2013). Assessing the moderating effect of consumer product knowledge and online shopping experience on using recommendation agents for customer loyalty. *Decision Support Systems*, 55(4), 883-893.
8. Hsu, M. H., Chang, C. M., & Chuang, L. W. (2015). Understanding the determinants of online repeat purchase intention and moderating role of habit: The case of online group-buying in Taiwan. *International Journal of Information Management*, 35(1), 45-56.
9. Heinrich, S. (2015). *Serious Business: How to attract and persuade customers without being salesy*. BoD–Books on Demand.
10. Palanivel, K., & Sivakumar, R. (2014). Development of Agent-based Personalized Recommender System for effective Human-Computer Interaction.
11. Wang, J. (2013). *Session aware recommender system in e-commerce* (Doctoral dissertation, University of California, Santa Cruz).
12. Aznoli, F., & Navimipour, N. J. (2017). Cloud services recommendation: Reviewing the recent advances and suggesting the future research directions. *Journal of Network and Computer Applications*, 77, 73-86.
13. Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: introduction and challenges. In *Recommender systems handbook* (pp. 1-34). Springer US.
14. Tejeda-Lorente, Á., Porcel, C., Peis, E., Sanz, R., & Herrera-Viedma, E. (2014). A quality based recommender system to disseminate information in a university digital library. *Information Sciences*, 261, 52-69.
15. Amatriain, X., & Pujol, J. M. (2015). Data mining methods for recommender systems. In *Recommender Systems Handbook* (pp. 227-262). Springer US.
16. Abbas, O. M., Mustafa, M. E., & Ibrahim, S. B. (2015). The Role of Data Mining in Information Security. *International Journal of Computer* (IJC), 17(1), 1-20.
17. Banaee, H., Ahmed, M. U., & Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors*, 13(12), 17472-17500.

ISSN: 2349-7637 (Online)