Big Data Analysis of Personalized Recommendation in E-Commerce

Xijun Lin

University College Dublin, Quinn Business School, Dublin 4, Dublin, Ireland
Email: xijun.lin@ucdconnect.ie

ABSTRACT
The personalized recommendation analysis for users' behaviors in e-commerce platform by big data analysis. Study on what kind of factors will affect on personal recommendations for goods on Amazon more effective by using Principal Component Analysis (PCA), Association Analysis and SVM (Super Vector Machine) Model. Using RFM Model to increase platform sales by analysis those data. This model aims to improve the accuracy and effectiveness of personalized recommendation, therefore improving the sales of goods on platform and the satisfaction of users.

Keywords: *Principal Component Analysis, Association Analysis, Super Vector Machine, E-commence, Users' Behavior*

1. INTRODUCTION

With more and more popular of the Internet, e-commerce platforms have gradually become one of the main shopping methods, for instance Amazon, Netflix, eBay etc. Those platforms will measure the personalized browsing behaviors by recommendation systems, some scholars believe that recommendation systems filter the key words of users' interactions and goods information in Recommender systems, which will suggest the product for customers by analyzing the K-nearest neighbor; Association and Clustering [1][2].

First, classify items via PCA to select the most purchased or the most hits' product. Moreover, association analysis is used to give goods and users' personal preferences to conduct specific analysis and classification. Finally, the same category (the same user preferences, commodity types) was classified by clustering, and the preferences were evaluated and reclassified. Compare the big data analysis with the reality situation, analyze the destabilizing factors, and consider how to improve the accuracy and make it closer to the reality. In this paper, it will use the PCA (Principal Component Analysis) and then by building the SVM Model (Supper Vector Machine Model) to classify the users' behavior which improves the accuracy of model and make it closer to the reality situation.

2. THEORETICAL BACKGROUND AND EXPECT DEVELOPMENT

In Recommender Systems announced the significance of personalized recommendation and it develop ability, profitability and effective [1]. With the advent of the 21st century, online shopping has become one of the mainstream shopping methods for public, and more and more e-commerce platforms have emerged. As of June 2021, the size of China's online shopping users reached 812 million, up 29.65 million from December 2020, accounting for 80.3% of overall Internet users [3].

In the value creation in e-business mentioned that efficiency, complementarities, lock-in, novelty are four dimensions of the value creation potential [4]. With the increasing of Internet users, the prospect of Internet development is getting brighter and brighter, and more physical stores or businesses are stationed on different e-commerce platforms to sell their goods. Therefore, e-commerce platforms will analyze the behavior of users in order to improve these four dimensions to growing the value of e-commerce platforms.

3. COLLECTING THE DIFFERENT KINDS OF DATA ON USERS

Aim to collecting the different types of data, in Aliyun Tianchi exhibits the millions of datasets, among which the dataset selected in this paper is called 'User Behavior Data from Taobao for Recommendation'. This

data is offered by Alibaba, it is including the user's implicit recommendation problem which shows by the 4 different types of behaviors of Taobao users are page view, purchase, collection and additional purchase. This paper will analyze the trends of the four behaviors by performing PCA (Principal Component Analysis), Association Analysis and SVM (Super Vector Machine), and finally the results will be analyzed and discussed with relevant theories.

4. DATA ANALYSIS OF PERSONALIZED RECOMMENDATION IN TAOBAO

4.1 Build The Model

First of all, this data analysis is based on Aliyun Tianchi dataset (user behavior dataset), which is from Taobao (one of the biggest e-commerce platform in China), using conversion funnel to analyze common e-commerce analysis indicators, including conversion rate, PV (Page view), UV (Unique visitor), repurchase rate, etc. During the analysis process, Python was used to Sorting out the data and conducting PCA. The cleaned data was imported into MySQL database, and the data was extracted using MySQL, then the data was analyzed using Association Analysis and SVM classification and prediction partly, and finally the results were visualized using Pyechart.

4.2 Analysis Of Page View

According to the analysis results of the following Figure 1 and Figure 2, the daily active time distribution and weekly active time distribution of users show that daily 18-22 hours and weekend are the peak active time for users to browse web page, so merchants should focus their resources on attracting traffic and marketing activities during this golden time, such as pushing users' favorite new products and best-selling products based on personalized recommendation algorithm during this active time. Since the dependency between location and promotion is not as strong on the Internet [5], most of the online marketing is done through advertising and live streaming with goods, which will be more effective to increase the purchase rate of users and sales of products. E-commerce platforms have to control the golden browsing time precisely, and significantly improve the means of advertising, marketing, promotion and sales promotion during this period, so that the benefits of the platform can be maximized.

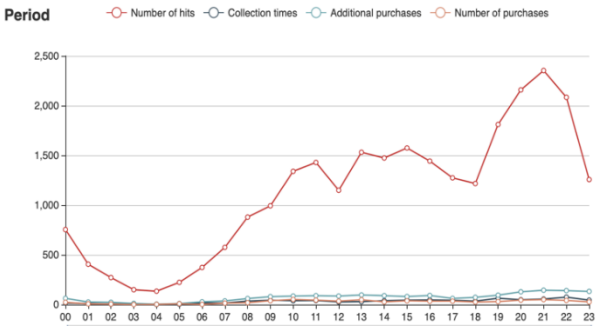

Figure 1 User daily active time period

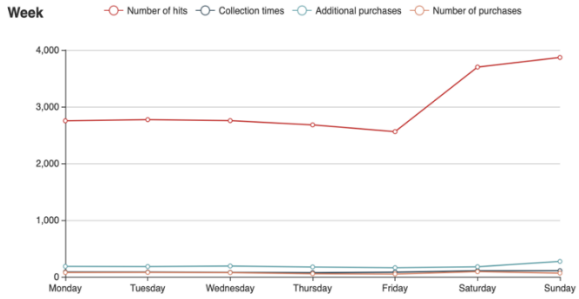

Figure 2 User weekly active time period

4.3 Analysis Of Repurchase Rate

According to the user repurchase rate calculation data below, Taobao repurchase rate of up to 63.77%, indicating that Taobao user quality and loyalty is still relatively high Figure 3, For repurchase users, we can combine RFM model [6] user value classification to conduct accurate marketing activities. On the other hand, for single-purchase customers, consider reminding customers of special offers in the form of customer service/text messages issued at regular intervals after their first purchase happens, in order to improve the repurchase rate. Merchants are able to analyze user data and then promote advertising to users who have purchased their products through advertising and other advertising methods.

Figure 3 Repurchase rate of users in Taobao

4.4 Analysis Of User Preferences

From the trend of 9 days shows on Figure 4, the conversion rate of the first 7 days is in a stable stage, indicating that the consumption situation after the double 11 tends to calm down, and after entering December, there is a slightly substantial increase in the browsing volume, a small increase in the conversion rate of adding purchases, and a significant decrease in the purchase conversion rate. Double 11 is the promotion activity festival of Chinese e-commerce platform. The price of goods on e-commerce platform will reach the lowest at that time, which is the best time for users to purchase goods, therefore it is normal to see that the purchase volume of goods decreases after Double 11.

Figure 4 Average daily visits on Taobao

5. RESULT AND DISCUSSION

The results of the products and categories that users may like are shown below Figure 5. Due to the user privacy clause, here both the products and categories can only show the ID, Name cannot be displayed to protect user privacy, otherwise the results could be more intuitive. Merchants can analyze the user behavior of the best-selling products to select products and thus optimize the product structure. In the analysis of the best-selling products, comparing the data from users' shopping paths to discover the differences between different products, and then find better optimization solutions to continuously optimize the products and improve the conversion rate.

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

CategoryID	count(BehaviorType)	Hits	Collection	Additional	Purchases
4756105	1630	1508	30	82	10
4145813	1192	1092	38	53	9
2355072	953	904	19	26	4
3607361	760	709	8	39	4
982926	755	685	14	52	4
4801426	664	598	19	41	6
1320293	561	513	15	29	4
2520377	556	511	21	24	0
2465336	476	443	11	21	1
3002561	458	408	20	23	7
4181361	406	364	10	28	4
149192	374	330	23	13	8

Figure 5 Hits ranking of products

Merchants can analyze the preferences of users in e-commerce platforms through RFM model [5]. RFM model can be used to identify customer segments for analysis and treatment, offered for accurate marketing activities. Due to the lack of price in the data, only the recency and the frequency be analyzed. In an instance, the number 4756105 category has the highest hits and additional in Figure 5, which mean it might be daily goods. However, the second category's collections times are higher than the number 4756105 category, second category's additional times are far less than the first one, indicating that this item may not be a necessity. Merchants can analyze the number of hits, collections, additional and purchases made by users on the products, and using the RFM model to make a simple review, expose the products with the highest recency and frequency at prime time.

E-commerce platforms are using data analytics for more effective and accurate promotions. Nowadays, people are living such a fast-paced life that e-commerce platforms not only enable merchants to profit from it, but by accurately analyzing user preferences, e-commerce platforms reduce the time users spend browsing products maximum, compared to some brick-and-mortar stores, the e-commerce platforms bringing people in a more convenient life where users can shopping at home. If e-commerce platforms are able to use Big data analysis (BDA) flexibly, it will facilitate universal use and rapid delivery of insights across organizations to maximize business value [7]. It also allows for better analysis and definition of the characteristics and types of big data in the future.

6. CONCLUSION

In this paper, using the PCA, Association Analysis and SVM model on the database, cleaning the data of consumers in e-commerce platform is analyzed in detail. With the advent of the Internet era, more and more merchants choose to promote, market, and sell their products through e-commerce platforms. This paper gives different feedback to products by analyzing RFM models of consumers' clicking, collecting, adding and buying behaviors, and merchants can optimize their products and improve customer conversions based on this data. Big data analytics allows companies and marketers to track and analyze consumer behavior and patterns, which in turn can have an impact on consumers' purchasing decisions [8]. E-commerce technology has had a significant impact not only on companies' export marketing, but also on their domestic platforms. A good e-commerce technology enhances sales support, thus improving distribution efficiency and promotional efficiency [9]. Therefore, the technology of e-commerce platforms needs to be continuously updated to bring more benefits to companies, merchants, and customers.

This paper enables e-commerce platforms and their merchants to gain more benefits from their customers by combining big data analysis and marketing strategies. At

the same time, e-commerce platforms can analyze user behavior along with user preference analysis to make more accurate product recommendations to users and make the platform sustainable. E-commerce platforms can update their algorithms for analyzing user preferences every now and then, which can improve the accuracy of individual user preference recommendations. By improving the efficiency, complementarity, targeting, and novelty of e-commerce platforms, e-commerce platforms can achieve longer-term growth and profitability. At the same time, e-commerce platforms can improve their service, relevance, and reliability, elements that are critical to merchants, businesses, and users [10].

REFERENCES

[1] Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58.

[2] Le, T. M., & Liaw, S. Y. (2017). Effects of pros and cons of applying big data analytics to consumers' responses in an e-commerce context. *Sustainability*, 9(5), 798.

[3] China Internet Network Information Center. (2021). *Statistical Report on China's Internet Development* [Ebook] (48th Edition., pp. 17). Retrieved from http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202109/P020210915523670981527.pdf

[4] Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic management journal*, 22(6-7), 493-520.

[5] Allen, E., & Fjermestad, J. (2001). E-commerce marketing strategies: an integrated framework and case analysis. *Logistics information management*.

[6] Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415-430.

[7] Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, 26(2), 173-194.

[8] Rosário A, Raimundo R. Consumer Marketing Strategy and E-Commerce in the Last Decade: A Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*. 2021; 16(7):3003-3024. https://doi.org/10.3390/jtaer16070164

[9] Gregory, G., Karavdic, M., & Zou, S. (2007). The effects of e-commerce drivers on export marketing strategy. *Journal of International Marketing*, 15(2), 30-57.

[10] Mohapatra, S. (2013). E-commerce Strategy. *In E-Commerce Strategy* (pp. 155-171). Springer, Boston, MA.