Construction of Social E-commerce Merchant Segmentation Model Based on Transaction Data

Haihui Xin[1*], Shanshan Zhang[2]
*Email: 1308649030@qq.com

[1] Dalian University of Science and Technology, Dalian, China

[2] Dalian Maritime University, Dalian, China

Abstract: With the rapid development of social e-commerce today, the statistical analysis of consumers on the platform by traditional e-commerce platforms is no longer suitable for the statistical analysis of user behaviour under the current social e-commerce. However, the different income levels of consumers and the different behaviours of using social software have put forward higher requirements for the marketing and promotion methods of social e-commerce. Therefore, it is necessary for social e-commerce to accurately subdivide merchants to identify their value, provide consumers with differentiated services, and implement more effective user strategies. In this paper, the indicators in the traditional RFM model are matched with the characteristics of social e-commerce merchants, the number of friends of social e-commerce merchants is introduced, and a RCFM model suitable for social e-commerce transaction data segmentation is constructed. In this paper, the weight of each index in the RCFM model is calculated by the analytic hierarchy process. Finally, the superiority of the new model in precise segmentation is verified through the weighted optimization and comparison experiment of the RCFM model. This research enriches the related research on social e-commerce business models, provides ideas for social e-commerce merchants' value judgment, and provides a foundation for social e-commerce enterprises to construct social e-commerce merchant portraits and implement targeted services.

Keywords: E-commerce information, Social value, Transaction data, Segmentation model

1 Introduction

In the iterative development from e-commerce to traffic e-commerce and then to social e-commerce, social e-commerce has become a hot topic in the e-commerce industry [1]. Under the background of the rapid development of mobile Internet, as a current fastest Social e-commerce, the most accurate, most effective, and the latest mobile new business model also faces severe challenges. User behaviour analysis based on social media content can help to understand consumers' consumption interests and fields, analyze the target user groups to which users belong, dig out the pain points of users and some problems existing in products, and at the same time, construct target user portraits according to users' behaviour habits and shopping habits [2]. Through data mining and extensive data analysis, the social e-commerce can know the matching relationship between users' consumption areas and enterprises' business areas; the social e-commerce can know the user's spending power; the social e-commerce can know the user's consumption preferences and the changing trend of

FFIT 2022, October 28-30, Shenzhen, People's Republic of China

DOI 10.4108/eai.28-10-2022.2328461

consumption preferences; the social e-commerce can calculate the probability of users buying enterprise products [3].

A recommender system for social networks based on the trust value is proposed by Walter et al. [4]. The trust value of the target participant is calculated by multiplying the trust value between the source node and the target node by the trust value of the participants on the shared path. A random walk model for both parties of online social network transactions is put forward by Tmstwalker et al. [5]. In the random walk model, the transaction buyer needs to perform a random walk with a fixed number of hops according to the seller's purchase path in the social network to get a score, and the score is used to evaluate the product.

The social e-commerce users are classified according to different attributes and dimensions, and it is necessary to analyze the multi-attribute characteristics of social e-commerce users. In reality, the base of social e-commerce users is huge. How to distinguish ordinary users from e-commerce users in social networks is a very important research topic. The e-commerce users are also ordinary social users at some time, and the social users can become e-commerce users at any time, which makes the classification of social e-commerce users more difficult. The social users can become the e-commerce users at any time, increasing the difficulty of classifying social e-commerce users. The e-commerce users are not uniformly registered, and the types of goods sold are different. It is challenging to classify social e-commerce users according to the available attributes displayed by users in social apps [7].

According to the collected data of mobile social e-commerce, how to conduct multi-dimensional modelling analysis of users; how to mine the areas that users are interested in and accurately mine the target population from social media; how to analyze social e-commerce users and social media content in a multi-dimensional way by using machine learning deep learning and other technologies is a problem that needs to be considered and solved.

Firstly, in this paper, according to the characteristics of social e-commerce merchants, the RFM model is improved and optimized, and the social e-commerce value model is explored by using the data in the social e-commerce warehouse. The advantages and disadvantages of the traditional RFM model and RCFM model with the number of friends attribute are compared, and the better RCFM model is decided as the social e-commerce value model. By using the analytic hierarchy process (AHP) to set the weight of the RCFM model, the accuracy of social e-commerce value segmentation can be improved so that it can reasonably reflect the value of social e-commerce; through the comparative experiment of the RCFM model weighted optimization, the superiority of RCFM model optimization is verified, and it shows that it has a good effect on the value segmentation of social e-commerce.

2 Data preparation and preprocessing

The primary experimental data in the study is the transaction data for the current month generated by 53,307 active software users in February 2019 collected by the software. They were referring to the operating mode of the company's software in order to reduce the waste of enterprise operating costs caused by too much attention and investment to the lost users. In the research, when segmenting the value of social e-commerce merchants, the span of one month is used as the segmented period, and only users who are active in the software during this period are considered. For active software users outside the statistical span [8]. use Spark to collect statistics on the relevant data in the Hive data warehouse, and calculate the time interval (R), the total number of transactions in the month (F), the average transaction amount (PM) and other attributes[9] .

Part of the obtained data of the traditional RFM model is shown in Table 1.

Table 1 Social e-commerce RFM model indicator data

R	F	PM
0	1034	5.2235
2	56	114.4255
1	114	110.1574
3	101	291.7357
4	8	113.9325
0	12	137.8354
6	65	2028.0255
5	40	137.1269
7	106	2.3747
10	26	165.6264

3 Social e-commerce merchant segmentation model

Since the dimensions of each data indicator in the model are different, the data values between different attributes are also very different. The results will be affected if the initial value is used directly without processing [10]. Therefore, the dispersion standardization of the initial data is performed, and the corresponding transformation function is:

$$y_i = x_i - \min / \max - \min, (1 \le i \le n) \quad (1)$$

Since the recent transaction interval R index in the model has a negative correlation with the value of social e-commerce, the smaller the R value, the greater the value of social e-commerce merchants, so the corresponding transformation function to standardize the R index is:

$$y_i = \max - x_i \; / \max - \min, (1 \le i \le n) \quad (2)$$

The normalized data are shown in Table 2.

Table 2 Standardized social e-commerce indicator data

R	F	PM	R1	F1	PM1
1	0.1249	0.0256	0	1033	5.2423
0.9665	0.0247	0.0458	1	42	114.4243
1	0.0256	0.0232	0	151	110.1022
1	0.0152	0.0225	0	109	291.7224
0.8436	0.0455	0.0358	2	7	113.9233
1	0.0247	0.0288	0	30	137.8256
0.5547	0.0024	0.0267	12	12	2028.02354
1	0.0025	0.0563	0	21	137.2552
1	0.0156	0.0014	0	103	2.3442
0.1256	0.0014	0.0001	24	11	165.1544

4 Calculation of RCFM indicator weights

The single-level structure model of the established social e-commerce merchant RCFM model is shown in Figure 1.

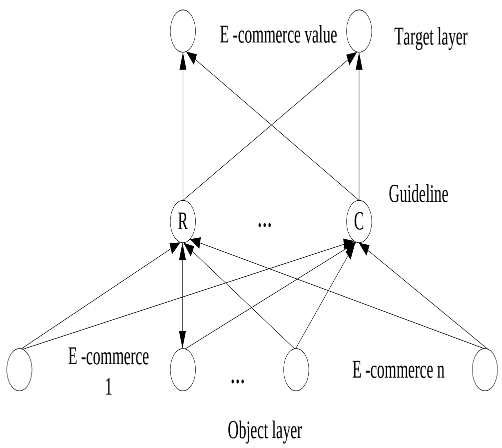

Figure 1 Single-level structure model of social e-commerce RCFM model

The four attributes that affect the value of social e-commerce merchants in the model are the recent transaction time interval R, the transaction frequency F, the average value of a single transaction M, and the number of social e-commerce merchant friends C. Comparing the influence degree of the four attributes on the value of social e-commerce merchants affects the weight setting of each attribute index of this layer on the value of upper-level social e-commerce merchants. The research objects involved in this research: Wechat merchants who conduct marketing on WeChat, based on considering the operation mode, sales method and profit point of such social e-commerce merchants, social e-commerce data analysts and senior operators of social e-commerce auxiliary software determined the corresponding weights of the four attributes in the model after repeated discussions and research.

In contrast, from the above analysis of the characteristics of social e-commerce merchants' operation and profitability, the single transaction amount of a social e-commerce merchant's transaction order is not very important to the value of social e-commerce merchants. The time interval between the last transaction and analysis point of social e-commerce merchants reflects the activity of social e-commerce merchants. As an industry with a high frequency of promotion and transactions, social e-commerce information on social networks is highly time-sensitive. Merchant's product promotion information also has such characteristics. Suppose social e-commerce merchants do not generate transaction orders for a long time. In that case, social e-commerce merchants may reduce the promotion of product information or the promotion effect is low and will negatively impact the value of the social e-commerce merchant during the period.

To sum up, for the RCFM model of social e-commerce merchants, the importance of the indicators is sorted, from high to low, C, F, R, M, and finally, the judgment matrix is obtained.

$$W = \begin{bmatrix} 1 & 1/2 & 1/3 & 5 \\ 5 & 2 & 4 & 7 \\ 4 & 3 & 1 & 6 \\ 1/2 & 2 & 1/3 & 1 \end{bmatrix}$$

Eigenvectors of a matrix $W = (0.12, 0.50, 0.33, 0.05)^T$.

Calculate the consistency index $CI = 0.079$. The corresponding consistency ratio CR value is 0.0881, which is lower than 0.1. Therefore, the above matrix meets the requirements of consistency, that is, the weights of R, C, F, M in the RCFM model R, C, F, M are 0.12, 0.50, 0.33, 0.05.

5 Experiment analysis

When the traditional RFM model is used for segmentation, it is generally considered that the importance of each index in the model is the same. However, the degree of impact of indicators on value will vary with the industry's development and the industry's characteristics. The weighted RCFM model is more in line with the characteristics of the social e-commerce industry. The research on merchant value focuses more on the promotion ability of social e-commerce merchants and the size of the spread of promotional information.

The optimal K value that needs to be set to obtain better results for RCFM model clustering is 3, and the K-Means algorithm optimized by Mean-shift is used to cluster the unweighted and weighted RCFM model data set with a K value of 3, and the index values corresponding to the results are shown in Table 3.

Table3 Weighted optimization comparison results

	Unweighted	Weighted
Davidson Boding Index DB	0.45552	0.4556
Silhouette coefficient SC	0.7458	0.7371

The comparison of the evaluation indicators in the results shows that the weighted RCFM model dataset has a better similarity of the same type and dissimilarity of different types than the unweighted result cluster.

From the results obtained after clustering, the ratio of the number of social e-commerce merchants in the resulting cluster clustered by the unweighted RCFM model is 28156:12755:12396, as shown in Figure 2.

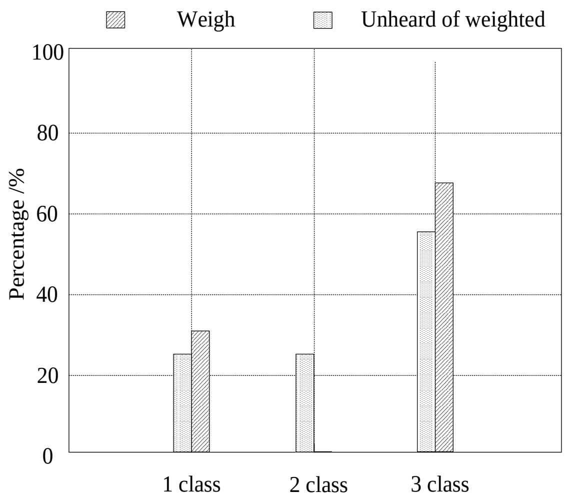

Figure 2 Clustering results of unweighted RCFM model

The average values of social e-commerce merchant attributes in the three categories are shown in Table 4.

Table 4 Average values of three types of social e-commerce attributes

	R	F	PM	C	USER
1 class	1.445	20.225	427.871	10.374	21257
2 class	9.914	7.524	349.364	4.091	12565
3 class	22.015	4.936	296.137	2.905	12357

From the average values of attributes in the three result clusters, it can be seen that although the unweighted RCFM model incorporates the features concerned by the social e-commerce industry into the value evaluation of social e-commerce, the importance of each index is regarded as the same, which leads to the gradual increase of only four attribute values among each class in the result of clustering the unweighted RCFM model data set. This clustering result is different from the company's actual situation of social e-commerce.

6 Conclusion

In order to achieve precise segmentation of social e-commerce merchants and provide differentiated services according to the characteristics of various social e-commerce merchants, the RFM model combined with social attributes is optimized in this paper and builds an RCFM model to reflect the value of social e-commerce merchants.

(1) Based on social e-commerce merchants' social attributes, some RFM model data is standardized in this paper.

(2) In this paper, AHP is used to optimize the index weight selection of the RCFM model so that the importance of each index in the model is consistent with the characteristics of social e-commerce merchants.

(3) A comparative experiment of weighted optimization of the RCFM model is carried out to verify the model's effectiveness for deep accurate segmentation.

This study only applied a machine learning algorithm for cluster analysis, there is still room for optimization at the level of the application algorithm, and the selection of evaluation indicators also has limitations. The follow-up work plans to introduce deep learning algorithms and more effective evaluation indicators to improve the segmentation accuracy of the model.

References

[1] Ángel Valarezo, Teodosio Pérez-Amaral, Teresa Garín-Muñoz, et al. Drivers and barriers tocross-border e-commerce: Evidence from Spanish individual behavior[J]. Telecommunications Policy, 2018, 42(6):464-473.

[2] Bin Hu, Qianqian Luo. Cross-border E-commerce Mode Based on Internet+ [J]. IOPConference Series: Materials Science and Engineering, 2018, 394(5)

[3] Johnson R, Zhang T. Deep pyramid convolutional neural networks for text categorization[C] 2017: 562-570.
[4] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations]!]. arXiv preprint arXiv: 1802.05365, 2018.
[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C], Advances in Neural Information Processing Systems, 2017: 5998-6008.
[6] Mohammad Malli, Nadine Said & Ahmad Fadlallah. A New Model for Rating Users* Profiles in Online Social Networks[J]. Computer & Information Science, 2017, 10(2):39-51.
[7] Chen W , Feng G , Zhang C , et al. Development and Application of Big Data Platform for Garlic Industiy Chain[J]. Computers, Materials and Continua, 58(1):229-24& 2019.
[8] Yin, Libo, et al. Relation Extraction for Massive News Texts. Tech Science Press, Computers, Materials and Continua, 60(1): 275-285. 2019.
[9] L. Wu, B. Yang, M. Jian5J. Sang D. Zhang and L. Zhang.MPPR:Multi Perspective Page Rank for User Influence Estimation.2018 IEEE International Conference on Big Data and Smart Computing (BigComp). 2018:25-29.
[10] A. Peleshchyshyn, V. Vus, O. Markovets and S. Albota.Identifying Specific Roles of Users of Social Networks and Their Influence Methods.IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT).2018:39-42.