International Journal of Product Development

ISSN online: 1741-8178 - ISSN print: 1477-9056
https://www.inderscience.com/ijpd

The deep mining of consumer behaviour data on product network marketing platform

Chunming Yu, Xin Jin

DOI: 10.1504/IJPD.2023.10058413

Article History:

Received:	02 April 2022
Last revised:	01 June 2022
Accepted:	03 October 2022
Published online:	05 April 2024

The deep mining of consumer behaviour data on product network marketing platform

Chunming Yu

Department of Information Management,
Shanghai Lixin University of Accounting and Finance,
Songjaing District, Shanghai, China
Email: 3387672@qq.com

Xin Jin*

Marketing Department,
AIWAYS Automobile Co., Ltd.,
Shanghai, China
Email: xinjin@mls.sinanet.com
*Corresponding author

Abstract: To overcome the problems of low accuracy of behaviour analysis results and low purchase proportion in traditional methods, a deep mining method of consumer behaviour data on product network marketing platform is proposed. Firstly, according to the nearest neighbour data distribution, the relevant subspace of consumer behaviour data of product e-marketing platform is divided, and the sparsity difference of relevant subspace data in each attribute is calculated. Then, the filtering of outlier interference data is completed by setting the difference threshold of local sparsity factor. Finally, the multi-source data mining method is used to analyse the difference between the data attribute weight and the target weight of each attribute behaviour, so as to realise the in-depth mining of consumer behaviour data. Test results show that the maximum error of consumer behaviour analysis of the design method is only 1, and the purchase proportion after secondary marketing reaches 10.2%.

Keywords: network marketing platform; consumer behaviour data; deep mining; nearest neighbour data; local sparse factor; outlier data.

Reference to this paper should be made as follows: Yu, C. and Jin, X. (2024) 'The deep mining of consumer behaviour data on product network marketing platform', *Int. J. Product Development*, Vol. 28, Nos. 1/2, pp.13–23.

Biographical notes: Chunming Yu received her PhD degree from Beijing University, China. Now, she teaches in Shanghai Lixin University of Accounting and Finance. Her research interests include e-marketing, electronic commerce and electronic payment system.

Xin Jin received his Master degree from Jilin University, China. Now, he works in AIWAYS Automobile Co., Ltd. His research interests include marketing strategy, sales marketing and brand strategy.

1 Introduction

The development of Internet technology not only changes people's work style and life style, but also makes some people's living habits develop in the direction of intelligence to a certain extent. Most obviously, the network-based shopping model has realised the coverage of the whole age group, the frequency of people's online shopping is increasing and the types of online shopping products also present diversified characteristics. In this context, e-commerce competition is becoming more and more intense, in order to improve product sales, professional product network marketing platform came into being (Abugu et al., 2020). To some extent, users' feedback on marketing information on the platform reflects their purchasing attitude to the product, so it is of great practical value to dig deeply into it to improve marketing effect.

In response to this, Huang and Jiao (2021) a method for in-depth data mining of consumer behaviour based on SRIMS and Holt winters. Taking the product online marketing platform as the research basis, SRIMS and Holt-Winters are used to analyse the consumer behaviour of users and predict their purchase behaviour. Experiments show that this method can effectively predict consumer behaviour, but this method is highly dependent on basic data, when interference information appears in the data, it will directly affect its data mining and prediction results. Cui and Wang (2020) proposed a method for deep mining of consumer behaviour data based on correlation analysis. Taking 'little red book' users as the research object, taking full account of the situational state corresponding to the behaviour of social e-commerce users, this paper analyses their consumption intention and purchase behaviour, and realises the accurate analysis of the correlation between consumers' purchase intention and behaviour, so as to mine consumer behaviour data. However, in practical application, it is found that this method has the problem of low accuracy of behaviour analysis results, and the actual application effect is not good. Liu and Zhang (2019) proposed a method for deep mining of consumer behaviour data based on SAS variable clustering method Through data pre-processing, de-empting, de duplication and stop words, using Chinese word segmentation and word frequency statistics of rostcm6 software, the comments are classified and analysed, differentiated selling point analysis, text analysis and emotion analysis and the emotion distribution map is drawn. The quantitative text is transformed into a matrix, and the SAS variable clustering method is used to deeply mine consumer behaviour data. However, when this method is applied to practice, it is found that this method has the problem of low purchase proportion, and the practical application effect is not good.

Owing the complexity of the consumer behaviour data of the marketing platform and the interference of various factors, it is difficult to analyse it, resulting in the problems of low accuracy of the behaviour analysis results and low purchase proportion in the traditional methods. This paper takes solving the problems of the traditional methods as the research goal, and puts forward a deep mining method of consumer behaviour data on product network marketing platform. Therefore, this method has the characteristics of high accuracy of consumer behaviour analysis results and high purchase proportion. Through the research of this paper, we also hope to provide a valuable reference for the intelligent development of product network marketing platform. The specific structure of the article is as follows:

1 The pre-processing of consumer behaviour data of product online marketing platform mainly includes the division of relevant subspace and the analysis of the divided subspace data. According to the difference of local sparsity factor, the sparsity difference of data in each attribute is calculated, which makes a basic basis for mining consumer behaviour data of product online marketing platform.

2 The filtering of outlier interference data is completed by setting the difference threshold of local sparsity factor. Multi source data mining is used to analyse the difference between the data attribute weight and the target weight of each attribute behaviour, so as to realise the in-depth mining of consumer behaviour data.

3 The experiment verifies the specific performance of the designed method, takes 300 users as the analysis object, carries on the actual test research and analyses the experimental results.

2 Pre-processing of consumer behaviour data for product online marketing platforms

2.1 Partition of related subspaces

Before the deep mining of consumer behaviour data of product network marketing platform, there may be more data clusters in the data set due to the influence of many different mechanisms (Dangl et al., 2020). It needs to be made clear that the attributes of each data cluster are different under different mechanisms, and attributes are one of the key factors that most directly reflect the local characteristics of data (Hosseinian and Butenko, 2021). Therefore, this paper firstly divides the consumer behaviour data of product network marketing platform into subspaces. Based on the local correlation shown by the attribute of data cluster, when the number of data subsets generated by a certain mechanism reaches a certain amount, the data of consumption behaviour is considered to have correlation (Smith et al., 2020). On this basis, the data related subspace division is mainly based on the distribution of outlier data in each data subset.

Assume D is a subset of consumption data containing n attribute dimensions, and there is a full-space $F = S$ for D, where S represents the attribute set of the data and exists

$$S = \{S_1, S_2, ..., S_n\} \tag{1}$$

where S_d represents the attributes of consumer behaviour data (Petcharat and Leelasantitham, 2020).

On this basis, arbitrary data x_i in the data is taken as an object, and the data of its nearest neighbour can be represented as

$$LD(x_i) = N(x_i, F) \tag{2}$$

The LD represents the nearest neighbour data of x_i, and N represents the position relation function. On this basis, when the distribution of $N(x_i, F)$ on the S_d attribute dimension is uniform, it can be considered that the S_d attribute does not have the function of feedback the value of consumer behaviour data, and this paper divides this

kind of data into unrelated subspaces. When $N(x_i, F)$ is non-uniform in the S_d attribute dimension, the S_d attribute has the function of feedback the value of consumer behaviour data. For this kind of data, this paper divides it into related subspaces.

In this way, we can avoid the difference of data set distribution caused by the adjustment of attribute dimension, reduce the number of failed data and improve the reliability of the final data mining results.

2.2 Subspace data analysis

After dividing the data-related subspace, it is necessary to analyse the data in the child control to determine the relationship between the data. For the nearest neighbour data of x_i, this paper uses the sparse difference of local data attribute as the basis to judge its distribution. In the Sub-section 2.1 part, we have only filtered the data with uniform distribution. Therefore, when we analyse the data of consumer behaviour with non-uniform distribution, we introduce coefficient factor (Dong et al., 2020). Assuming x_i is the *i*-th data object in data set D, $LD(x_i)$ is a local data set consisting of x_i and its nearest neighbour. Then in $LD(x_i)$, the degree of local sparsity of x_i on each attribute can be expressed as

$$z(x_i) = \frac{x_i \to S}{N(x_i, F) \to S} \tag{3}$$

where $z(x_i)$ indicates the degree of local sparsity of x_i on attribute S.

From the formula (3), it can be seen that the greater the local sparse factor $z(x_i)$, the more sparse the distribution of $LD(x_i)$ on attribute dimension S, and conversely, the smaller the local sparse factor $z(x_i)$, the more dense the distribution of $LD(x_i)$ on attribute dimension S.

In this way, the local sparse factor in attribute space of all data subsets of consumer behaviour data can be calculated. In this case, the sparse factor matrix of data set D, which can be expressed as

$$[B] = [Z_n] \tag{4}$$

Among them, $[B]$ represents the sparse factor matrix of consumer behaviour data set D, and $[Z_n]$ represents the local sparse factor of data in n attribute dimensions.

On this basis, we can realise the calculation of the difference of sparsity degree on each attribute according to the difference of local sparse factors, which can be expressed as

$$\varepsilon = \sqrt{\left[\frac{(Z_n - Z_{n1})}{Z_n}\right]^2} \tag{5}$$

Among them, ε means data in each attribute sparse degree difference. Through formula (5), it can be seen that the larger the value of ε is, the greater the difference of

local sparse factor matrix of data object x_i is, the more uneven the distribution density on attribute dimension is. On the contrary, the smaller the value of ε is, the smaller the difference of local sparse factor matrix of data object x_i is and the more uniform the distribution density on attribute dimension is.

2.3 Local outlier data filtering

Through the analysis of the data in the Sub-section 2.2 part, we can find that the data in the relevant subspace still exist in the form of outliers, so we need to filter out the data to improve the accuracy of the final mining results.

Firstly, the threshold of local sparse factor difference is set (Araki and Akaho, 2020). Considering that there are many attribute dimensions of consumption behaviour data and the overall scale of data is relatively large, this paper sets the threshold of local sparse factor difference based on the total amount of data in subspace and the actual distribution of data. Assuming that the position of x_i in its corresponding correlation sub-space is the centre, the upper limit of the local sparse factor difference of the data subset can be expressed as $z(x_i)$

$$\max \varepsilon = \frac{al * z(x_i)}{(1-\lambda)} \tag{6}$$

Among them, $\max \varepsilon$ represents the upper limit of the local sparse factor difference of the data subset of x_i, a represents the correlation coefficient of the data subset of x_i, l represents the maximum distance between x_i and the data subset of x_i and λ represents the uniformity coefficient. In Formula (6), the value of l depends mainly on the distribution of the data. In this paper, we set the value of l according to the data standard of 80% or more of the coverage data subset.

$$l = \frac{0.8k}{\lambda} \tag{7}$$

where k represents the total amount of data in the subset of x_i. The upper limit of the local sparse factor difference for a subset of data is calculated as follows:

$$\max \varepsilon = \frac{0.8ak * z(x_i)}{\lambda - \lambda^2} \tag{8}$$

Considering that a large number of failed data may lead to the final data subset of data is not representative, the mining of it will have no practical value. Therefore, for the set of the lower limit of the local sparse factor difference of the data subset, this paper also sets it according to the coverage level of the data subset above 60%.

In order to improve the value of data mining, the data outside the threshold is filtered.

3 Deep mining of consumer behaviour data

After filtering the local outlier data, reliable consumer behaviour data will be obtained. Combined with the types of consumer behaviour data on the product online marketing

platform, this paper divides the data depth mining into two stages. First, the data weight is determined. According to the determination results, the multi-source data mining method (Korhonen et al., 2021) is used to analyse the behaviour impact caused by the difference between the data attribute weight and the target weight that generates each attribute behaviour, so as to realise the mining of behaviour data.

3.1 Data attribute weight determination

Weights of data attributes are determined on the basis of product marketing plans (Jin et al., 2021). To this end, first of all, the marketing plan is partitioned, so that a complete marketing data is divided into multiple individuals, so that data mining can be carried out from multiple starting points simultaneously. Based on all the pre-processed data, this paper divides them into several types.

When calculating the weights of data attributes, first of all, the data scale and marketing scope shall be calculated, then the frequency of marketing information promotion shall be calculated and finally the attributes of data feedback targets shall be determined (Rezaei et al., 2021). According to the calculation results of the marketing plan, check the data in the behaviour of each attribute, assume the marketing technology is f, the corresponding product information promotion frequency is t and the number of consumers covered by marketing is c. The weight of each attribute of pre-processed data can be expressed as follows:

$$w(n) = \frac{ftc}{n} \tag{9}$$

where $w(n)$ represents the weight of the n attribute. In the actual calculation process, the actual implementation of the marketing plan may be different from the planning stage, so it is required to ensure that the parameters are set on the basis of the actual data.

Through this way, we can set the weight of data attribute, and provide basic guarantee for consumer's behaviour data mining.

3.2 Behavioural data mining

After the weights of the attributes of the actual data are determined, the behavioural data of the attributes are analysed and calculated by analysing the differences between the weights of the attributes and the target weights that generate the behaviours of the attributes (Kiaei and Matin, 2022).

If the difference between the weights of the data indicators and the target weights of the behaviours that generate the attributes is e, then the actual marketing data will differ accordingly. This difference can be expressed as

$$\Delta = \frac{eD}{T} \tag{10}$$

Among them, Δ indicates the degree of difference between the attribute weights of the actual marketing data and the weight of the behaviour objectives of the attributes, and T indicates the implementation time of the marketing plan, when it is calculated on the basis of the frequency of promotion information release, which can be expressed as follows:

$$T = mt \tag{11}$$

Among them, m indicates the period multiple of frequency of promotion information release. Based on this, using multi-source data mining (Fallmann et al., 2020), it can be found that the time to implement the marketing plan for changes in consumer behaviour data is mainly reflected in two forms: the number of people covered and the number of people browsed. The change in the number of people covered can be expressed as

$$\Delta_r = tq \tag{12}$$

Among them, Δ_r means to cover the number of changes in q means a single marketing coverage.

And the number of browsing changes can be represented as

$$\Delta_l = tQ \tag{13}$$

Among them, Δ_l represents browsing times change circumstance, Q represents the browsing times of single promotion.

Each attribute behaviour is greatly affected by the local economic environment, so it is necessary to calculate the difference between the number of people covered by marketing and the target weight of each attribute behaviour based on the average salary of the region, which can be expressed as follows:

$$\Delta_{dr} = T\left(q - q'\right) \tag{14}$$

Among them, Δ_{dr} represents the impact of local economy on the number of people covered by each attribute, q' represents the number of people covered by actual single marketing and then combined with data size, the number of people producing each attribute is

$$P_r = \frac{\left(T - t\right)q'}{w\left(n\right)} \tag{15}$$

Among them, P_r represents the number of people who produce the behaviour of each attribute, through this way, the behaviour information is mined to complete the data mining.

4 Test analysis

In order to further analyse the effect of some designed deep mining methods of consumer behaviour data, experiments are conducted.

4.1 Test data preparation

In this paper, a product network marketing platform for testing, randomly selected 300 users as the test object. In order to improve the reliability of test and the representativeness of test results, test objects will be selected from different shopping channels. The basic information of the test subjects was obtained from the backstage of the platform, including sex, age, city where they are located, monthly disposable income,

time of contact with the internet and frequency of online shopping. The data were statistically analysed by SPSS19.0 software, and the specific results were shown in Table 1.

Table 1 Basic statistics for test objects

Basic information	*Classification*	*Percentage/%*
Gender	Man	35.46
	Lady	64.54
Age/year	18–22	22.16
	22–27	36.41
	27–32	19.27
	32–40	20.10
	More than 40	2.06
City of Residence	First-tier city	10.85
	Second-tier city	25.46
	Third-tier city	30.22
	Fourth-tier cities	12.60
	Five-tier cities	11.45
	County-level cities	9.42
Monthly disposable income/$	1000–2000	30.35
	2000–3000	36.92
	3000–5000	16.44
	5000–10,000	10.05
	More than 10,000	6.24
Internet Access Time/Year	Within 1 year	20.60
	1–3	25.98
	3–5	30.44
	More than five years	22.98
Online purchase frequency/times/month	3–7	38.57
	7–10	22.10
	10–20	10.23
	20 +	2.25

Integrate, de duplicate and normalise the collected data to ensure the consistency of sample dimensions of experimental data. Take the processed data as the experimental sample data, divide the experimental sample data into test set and experimental set, input the test set into the simulation platform, obtain the optimal operating parameters of the platform and take the optimal operating parameters as the initial simulation parameters, so as to ensure the authenticity and reliability of the simulation results.

4.2 Test methods

On the basis of the above, the same online marketing information was respectively sent to the above-mentioned 300 users, with a total of 10 items in the form of short videos and the behaviour of statistical testing on marketing information by using SPSS19.0 software was the same. The specific results are shown in Table 2.

Table 2 Test object behaviour statistics

Behaviour	*Number/person*	*Percentage/%*
Ignore marketing information	46	15.33
Partially view marketing information	85	28.33
View complete marketing information	101	33.67
Like marketing information content	26	8.67
Message marketing information content	15	5.00
Browse the marketing message board	10	3.33
View marketing information repeatedly	12	4.00
Search shopping platforms for products	5	1.67

On this basis, the basic statistical information of the test objects in different behaviour groups is matched and the corresponding relationship between the behaviour and the user information is obtained.

4.3 Test results

After getting the behaviour information of the test subjects with corresponding relationship, this paper presents a data mining method to analyse the purchase behaviour of consumers by using this design method and Huang and Jiao (2021) method and Cui and Wang (2020) method, and compares the results with the actual results. Among them, the mining results for the above 8 behaviours are shown in Table 3.

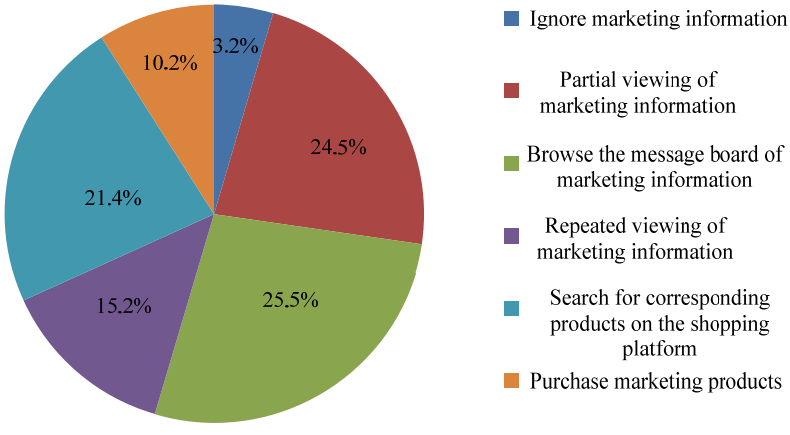
Table 3 Statistical results of consumer behaviour analysis by different methods

Behaviour	*Actual number/person purchased*	*Huang and Jiao (2021) method*	*Cui and Wang (2020) method*	*Method in this paper*
Ignore marketing information	1	0	0	1
Partially view marketing information	3	2	2	4
View complete marketing information	10	15	14	11
Like marketing information content	4	5	6	4
Message marketing information content	3	5	5	2
Browse the marketing message board	2	4	4	2
View marketing information repeatedly	3	5	6	4
Search shopping platforms for products	2	3	3	2

From the Table 3, we can see that the data mining method proposed in this paper is more accurate in analysing consumer behaviour. Among the analysis results of Huang and Jiao (2021) method and Cui and Wang (2020)method, the analysis error of the consumers who show the behaviour of completely viewing marketing information is the most obvious and the maximum error of the analysis results of the methods of this paper is only one person, and the analysis results of the consumers who show the behaviour of ignoring marketing information, clicking on the content of marketing information, browsing the message board of marketing information and searching corresponding products on the shopping platform are completely consistent with the actual results.

Based on the above, in order to further test the application effect of the data mining method designed in this paper, we analyse the test groups that show the behaviour of marketing information content, and analyse the reasons for the basic information of unpurchased consumers. Among them, there are 22 people. The feedback behaviour of this paper is shown in Figure 1, which is about the online marketing information of the groups with disposable income of more than 2000 yuan and online shopping frequency of more than 3 times per month.

Figure 1 Statistic chart of personalised secondary marketing effect (see online version for colours)



From the Figure 1, we can see that the proportion of purchasing behaviour after the second marketing is 10.2%, which shows that the analysis results of this paper are reliable and can provide valuable reference for the actual marketing activities.

5 Conclusions

1 The development of network has changed people's consumption habits and consumption patterns to a great extent. Under this background, the online product marketing platform aiming at product sales came into being, and the behaviour of users on the platform directly or indirectly fed back the purchase intention of users. In-depth analysis of it is of great value to improve the marketing effect.

2 Therefore, this paper proposes a deep mining method of consumer behaviour data on product network marketing platform, integrates multiple factors affecting consumer behaviour, realises the deep mining of behaviour data value and provides reliable guidance for the actual marketing work.

3 The experimental results show that the maximum error of consumer behaviour analysis of the design method is only 1, and the proportion of purchase behaviour after secondary marketing reaches 10.2%. Therefore, this method has the characteristics of high accuracy of consumer behaviour analysis results and high purchase proportion, and can be further popularised in practice.

References

Abugu, J.O., Chukwu, A.M. and Obasi, O.K. (2020) 'Exploring social media marketing in business: influence on product adoption perspective', *European Journal of Business and Management*, Vol. 12, No. 21, pp.2222–2839.

Araki, T. and Akaho, S. (2020) 'Spatially multiscale dynamic factor modeling via sparse estimation', *International Journal of Mathematics for Industry*, Vol. 11, No. 1, pp.1950005–1950016.

Cui, Q. and Wang, Y.R. (2020) 'Research on consumption intention and purchasing behavior of social e-commerce users under multi-dimensional situations – analysis of 'Xiaohongshu' users as data collection objects', *Price: Theory and Practice*, Vol. 12, No. 1, pp.95–98,163.

Dangl, R., Leisch, F. and Steinley, D. (2020) 'Effects of resampling in determining the number of clusters in a data set', *Journal of Classification*, Vol. 37, No. 3, pp.558–583.

Dong, C., Gao, J. and Peng, B. (2020) 'Varying-coefficient panel data models with nonstationarity and partially observed factor structure', *Heliyon*, Vol. 39, No. 3, pp.700–711.

Fallmann, S., Chen, L. and Chen, F. (2020) 'Enhanced multi-source data analysis for personalized sleep-wake pattern recognition and sleep parameter extraction', *Personal and Ubiquitous Computing*, Vol. 17, No. 3, pp.1–21.

Hosseinian, S. and Butenko, S. (2021) 'Polyhedral properties of the induced cluster subgraphs – ScienceDirect', *Discrete Applied Mathematics*, Vol. 297, No. 15, pp.80–96.

Huang, T.W. and Jiao, F. (2021) 'Consumption behavior forecast based on ARIMA and holt-winters', *Computer Systems and Applications*, Vol. 30, No. 8, pp.300–304.

Jin, X., Ji, Y. and Qu, S. (2021) 'Minimum cost strategic weight assignment for multiple attribute decision-making problem using robust optimization approach', *Computational and Applied Mathematics*, Vol. 40, No. 6, pp.1–24.

Kiaei, H. and Matin, R.K. (2022) 'New common set of weights method in black-box and two-stage data envelopment analysis', *Annals of Operations Research*, Vol. 309, No. 1, pp.143–162.

Korhonen, P.J., Wallenius, J., Genc, T. and Xu, P. (2021) 'On rational behavior in multi-attribute riskless choice', *European Journal of Operational Research*, Vol. 228, No. 1, pp.331–342.

Liu, N.N. and Zhang, Q. (2019) 'Analysis of consumer demand and product data mining technology based on e-commerce platform', *Statistics of Inner Mongolia*, Vol. 15, No. 1, pp.38–41.

Petcharat, T. and Leelasantitham, A. (2020) 'A retentive consumer behavior assessment model of the online purchase decision-making process', *Heliyon*, Vol. 7, No. 10, pp.e08169–e08188.

Rezaei, J., Arab, A. and Mehregan, M. (2021) 'Equalizing bias in eliciting attribute weights in multiattribute decision-making: experimental research', *Journal of Behavioral Decision Making*, Vol. 35, No. 2, pp.e2262–e2275.

Smith, J.S., Zubatyuk, R., Nebgen, B., Lubbers, N. and Tretiak, S. (2020) 'The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules', *Scientific Data*, Vol. 7, No. 1, pp.134–144.