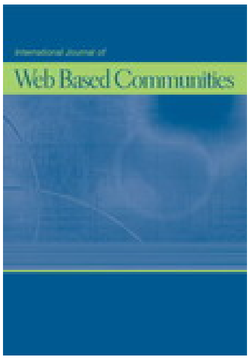


International Journal of Web Based Communities

ISSN online: 1741-8216 - ISSN print: 1477-8394
https://www.inderscience.com/ijwbc

A rough set-based consumer buying behaviour prediction method in online marketing system

Dian Jia

DOI: 10.1504/IJWBC.2023.10048316

Article History:

Received:	07 June 2021
Accepted:	04 January 2022
Published online:	20 January 2023

A rough set-based consumer buying behaviour prediction method in online marketing system

Dian Jia

School of Economics and Management,
Qinghai Nationalities University,
Xining 810007, China
Email: dianjia@36haojie.com

Abstract: Aiming at the problems of large prediction deviation and low acquisition accuracy of consumer purchase behaviour in traditional online marketing systems, a rough set-based consumer purchase behaviour prediction method in online marketing system is proposed. By improving the accuracy and recall rate of online consumer buying behaviour prediction methods, the deviation of prediction results is reduced. The data of consumer purchase behaviour in the region related to rough set are reduced to improve the accuracy and recall rate, and the forecast bias is reduced by removing redundant features in the e-marketing system. With the rough set theory, the dimension of consumer behaviour vector is reduced, and a predictive model framework is built. The simulation results show that the accuracy and recall rate of this proposed method are higher than 95%, and the minimum deviation of the prediction result is only 8.12%, which proves that the prediction result is more reliable.

Keywords: rough set theory; online marketing system; consumers; buying behaviour prediction.

Reference to this paper should be made as follows: Jia, D. (2023) 'A rough set-based consumer buying behaviour prediction method in online marketing system', *Int. J. Web Based Communities*, Vol. 19, No. 1, pp.64–77.

Biographical notes: Dian Jia received his PhD from Lyceum-Northwestern University, Republic of the Philippines. Currently, he works in School of Economics and Management, Qinghai Nationalities University. His research area includes business administration, human resource management and big data analysis, and marketing management.

1 Introduction

With the rapid development of the internet, consumers can obtain more resources and information, and businesses can use richer marketing means and marketing models. At the same time, it is more and more difficult for consumers to identify and screen information. How businesses can stand out in diversified marketing and attract more consumers has become the main problem at present. Consumer behaviour is not only a dynamic influencing factor, but also an influential changing factor. In the process of forming consumer purchase behaviour, there are external stimulating factors such as merchant marketing and post purchase evaluation of other consumers. The era of 'big data' continues to promote the generation of these stimulating factors. The stimulation of these stimuli in the 'big data' era includes merchant marketing and consumer post purchase evaluation. Compared with the traditional physical shopping behaviour, the shopping behaviour in the online marketing system is unstable, and consumers are more inclined to choose independently (Liu et al., 2020). Therefore, e-commerce enterprises need to more reasonably and scientifically mine consumers' purchase characteristics, accurately predict consumers' purchase behaviour, and thus formulate more targeted marketing methods (Lee and Lim, 2020). Consumer purchase behaviour prediction refers to the prediction of consumers' consumption ability and consumption level through questionnaire survey, so as to provide a certain theoretical basis for product R&D and marketing strategy formulation in the target market (Gillani et al., 2019). Therefore, it is of great significance to design an effective network marketing system to predict consumers' purchase behaviour (Davydenko and Peetz, 2020). Relevant scholars have also done a lot of research on it.

A reference-based method for predicting consumer behaviour (Ge et al., 2019) is proposed. This method reduces the attributes of consumption behaviour through difference matrix, and predicts consumers' purchase behaviour based on an attribute set, which has a good prediction effect. However, this method has the problems of low accuracy and low recall. Song and Tang (2019) propose a consumer psychological behaviour reflection and prediction method based on SOR model. The model studies consumers' shopping information clues, including price and perceived value, analyses the influencing factors of consumers' emotion, and thus judges and analyses the behaviour affecting consumption intention. Using the method of questionnaire, this paper makes an empirical analysis and research. However, this method has the problem of large deviation of prediction results. Yong (2019) proposes a method for predicting consumers' purchase intention based on Markov chain. This method analyses the impact of product promotion behaviour on purchase intention, takes skin care products as the experimental object, simulates promotion activities, analyses the market changes before and after promotion, and calculates the market share. The results show that promotional activities play a certain role in improving consumers' purchase intention. However, this method has the problem of low recall.

In order to solve the problems of large prediction deviation and low acquisition accuracy of consumer purchase behaviour, this paper proposes and designs a new prediction method of consumer purchase behaviour in network marketing system. The technical route of this paper is as follows:

1. In order to improve the prediction effect of consumer purchase behaviour, the accuracy and recall rate of online consumer purchase behaviour prediction methods are improved to improve the prediction accuracy of consumer behaviour. The introduction of rough set theory has unique advantages in dealing with uncertain and inconsistent information. In the upper and lower approximation domain, boundary domain and negative domain of rough set, the amount of data of consumer purchase behaviour can be reduced by control, and the accuracy and recall of prediction methods can thus be fundamentally reduced to the amount of data of consumer purchase behaviour.

2. It is based on improving the prediction accuracy of consumers' purchase behaviour and reducing the prediction deviation. By eliminating the redundant features in the E-marketing system, the prediction deviation is reduced. Rough set theory is used to

reduce the dimension of consumer behaviour vector, extract consumer purchase behaviour indicators, and build a prediction model framework.

3 With the prediction accuracy and prediction deviation of consumers' purchase behaviour in the network marketing system as the comparison index, experimental verification is carried out to compare the proposed method with methods of Ge et al. (2019), Song and Tang (2019) and Yong (2019).

2 Rough set theory is used to deal with consumer purchasing behaviour data

Rough set theory has unique advantages for the processing of uncertain and inconsistent information. Therefore, this study can predict whether customers' purchasing behaviour will occur in the future by reducing the data of consumers' purchasing behaviour.

In this study, the data of consumer purchasing behaviour is reduced by controlling the upper and lower approximate domain, boundary domain and negative region of rough set, so as to improve the accuracy and recall of the prediction method. Then on the basis of the decision attribute set and the conditional attribute set, the prediction bias is reduced by removing the redundant features of the online marketing system.

For any kind of conditional attribute $k \in K$ of consumption information, there exists a function k: $K \rightarrow T_k$, and T_k is represented as the range of the attribute k. Each element in W is called an individual, an object, or a row.

For any given subset of the star $I \subseteq K$ and for any given individual $x \subseteq K$, the following information functions are satisfied:

$$\ln f_I(x) = \{(k.k(x)) : \mathrm{k} \in I\} \tag{1}$$

where the $k(x)$ represents the data node. According to the above-mentioned results, the network node concept $\mathrm{X} \subseteq W$ and the attribute subset $I \subseteq K$ were the same, then the upper and lower approximation domain definitions for concept X were obtained:

$$\begin{cases} \overline{I}(X) = \overline{I}(IS, X) = \{x \mid (x \in W) \wedge ([x]_I \cap X \neq \varnothing)\} \\ \underline{I}(X) = \underline{I}(IS, X) = \{x \mid (x \in W) \wedge ([x]_I \cap X)\} \end{cases} \tag{2}$$

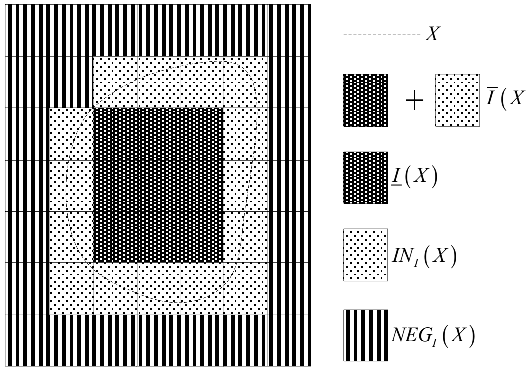
where $(x \in W)$ states the prediction parameters. The set $IN_I(X) = \overline{I}(X) - \underline{I}(X)$ denotes the I boundary region of $X; POS_I(X) = \underline{I}(X)$ denotes the I positive region of X; $NEG_I(X) = W - \overline{I}(X)$ is the I negative region of X (Toussaint et al., 2020). Schematic diagrams of upper and lower approximate domains, boundary domains and negative regions are shown in Figure 1.

The parameters in Figure 1 are used to reduce the data attributes of consumers buying behaviour in online marketing system. Under the condition that the classification ability of educational network data remains unchanged, attribute reduction is carried out, and redundant attributes are excluded. Given a database $M = (W, S)$ and an equivalent relation family $G \in S$ in the database, for any $H \in G$, if the following equation exists:

$$IND(G) = IDN(G - \{H\}) \tag{3}$$

Then, we call the online education network data H unnecessary in G, otherwise, we call the data H necessary in G. If for each $H \in G$, H is necessary in G, then G is called independent; otherwise, G is called dependent or not independent, so as to achieve attribute reduction of online education network nodes (Dong and Jiang, 2019).

Figure 1 Schematic diagram of upper and lower approximate domain, boundary domain and negative region



Rough set is a method to deal with uncertainty problems, so the uncertainty index measurement is carried out to deal with the uncertainty of consumer purchasing behaviour data in online marketing system. It was known that in the online marketing system $IS = (W, K)$ with a condition $X \subseteq W$, the supremum subset $I \subseteq K$ with respect to the concept $x \in W$ were the same, and the way of membership calculation was expressed as follows:

$$\varphi(I, X, x) = \frac{\left|[x]_I \cap X\right|}{\left|[x]_I\right|} \tag{4}$$

where | | is the potential of set. It was assumed that in the marketing system $DS = (W, K, a)$, there was a conceptual condition $X \subseteq W$ n, and the decision attribute a depended on the condition attribute set K with degree $h(0 \leq h \leq 1)$, then there was:

$$h = \lambda(K, a) = \frac{\left|POS_K(a)\right|}{|W|} \tag{5}$$

where λ represents the degree of dependence. According to this formula and formula (4), the uncertainty index can be obtained as follows:

$$\beta = (I, X)1 - \frac{\varphi}{\ln h} \times \left|\kappa \cdot \frac{\overline{I}(X)}{\underline{I}(X)}\right| \tag{6}$$

where κ represents the influence factor of uncertainty degree index; β is the uncertainty index of online marketing system (Das and Ramalingam, 2019). Based on fuzzy rough set theory, the online marketing system attributes are eliminated and reduced. In view of the uncertainty of marketing system, the uncertainty index is measured. Rough set attribute reduction can be carried out as follows:

Step 1 Set the decision attribute set and condition attributes set to D and C, respectively, calculate the conditional entropy $H(D|C)$, get global condition entropy. It belongs to the decision table T.

Step 2 Calculate the kernel attribute set $C0$ of D in C, and place the attribute set that does not belong to the kernel condition in $Att(Att = C - C0)$.

Step 3 Let $B = C0$. If $|B|! = 0$, then conduct the calculation to obtain conditional entropy $H(D|B)$, jump straight to process 6.

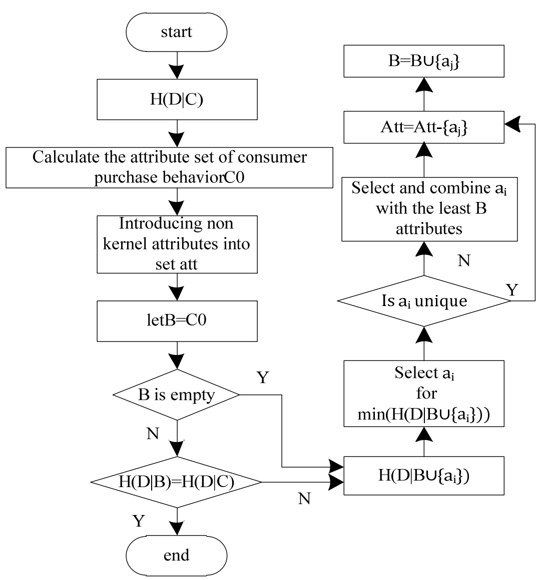
Step 4 Calculate decision set D of different attributes of $a_i \in Att$ C, and calculate the corresponding condition attribute set $B \cup \{a_i\}$ condition entropy $(H(D \mid B \cup \{a_i\})$.

Step 5 In all conditional entropy $H(D \mid B \cup \{a_i\}$, the minimum entropy value attribute a_j is selected. If the minimum value with multiple and value are consistent, then with the reduction of attribute set the number of attributes combination B minimum entropy is selected. Delete the selected attribute $(Att = Att - \{a_i\})$ in Att and introduce it into B, $B = B \cup \{a_i\}$.

Step 6 If there is a $H(D \mid B) = H(D \mid C)$, then B is changed into calculation results for the reduction of collection, otherwise jump back to the Process 2.

Based on the above analysis, the attribute reduction process of rough set is shown in Figure 2.

Figure 2 Rough set attribute reduction process



To sum up, the rough set theory is introduced. According to the rough set upper and lower approximate domain, boundary domain and negative region control, the consumer purchasing behaviour data is reduced, the decision attribute set and the conditional attribute set are set, and the redundant features of the online marketing system are removed.

3 Prediction of consumer purchasing behaviour in online marketing system based on rough set

The prediction of consumer buying behaviour means that through the induction and analysis of the known customer information data, there are data omissions and acquisition errors in the acquisition process of customer data. When the customer features are the same, there will be differences in purchasing behaviour, which leads to the wrong prediction of customer buying behaviour. Rough set theory has a unique advantage for the processing of such uncertain and inconsistent information. It can reduce consumer data information and predict whether customers will purchase in the future.

The regression function in rough set theory calculation method is as follows:

$$\mathrm{f}(x) = Y^T \left(K + \lambda I_n\right)^{-1} k(x) \tag{7}$$

where $Y^T = s$; K represents a Gram matrix, $K = k(x_i, x_j)$; x_i and x_j are arbitrary two vectors in the number S; I_n represents the identity matrix; λ represents the coefficient of the identity matrix, and $k(x)$ represents the kernel transformation vector of all vectors in x and S. It is expressed based on Gaussian kernel function, and its formula is:

$$k_\sigma\left(x, x_i\right) = \exp\left(-\frac{\|x - x_i\|_2^2}{2\sigma^2}\right) \tag{8}$$

The above process can complete the single dimensionality reduction of consumer behaviour vector and eliminate the whole dimensionality reduction process, thus effectively reducing the amount of computation.

Before calculating the index weight, the data of consumer purchasing behaviour should be sorted out first. Since different consumer data have great differences, the consumer purchasing behaviour data is processed in a standardised way (Lee et al., 2019). Through KMO and Bartlett's test, KMO = 0.661 > 0.5, and significance Sig. of Bartlett's test SIG = 0.00 < 0.05. Therefore, consumption behaviour data are suitable for prediction index analysis.

The extraction and analysis results of consumer purchasing behaviour indicators are shown in Table 1.

Table 1 Extraction analysis results of consumer purchasing behaviour indicators

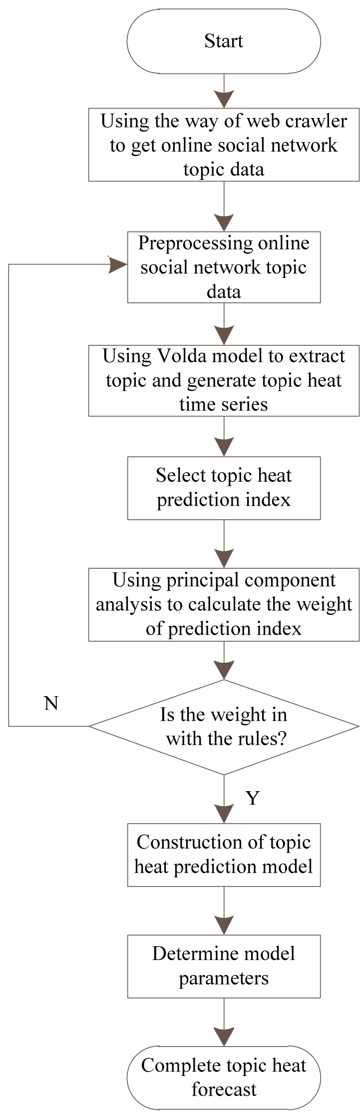
Principal component	*Variance contribution rate*	*Cumulative contribution rate*	*Characteristic value*
F1 Click (PV)	29.49	29.49	4.72
F2 Purchase (BUY)	21.98	51.47	3.52
F3 Add to cart (CART)	8.88	60.35	1.42
F4 Like (FAV)	4.76	85.26	0.76

In selection of indicators of consumer buying behaviour, the following rules should be followed: eigenvalue >1 or variance contribution rate >0.85. Table 2 is the index initial factor load matrix.

Table 2 Initial factor load matrix table

Principal component index		*F1*	*F2*	*F3*	*F4*
C1	Topic manager component	0.14	0.41	0.57	0.06
C2	Topic manager visibility	–0.21	0.78	0.27	–0.37
C3	Topic manager activity	–0.08	0.06	0.40	–0.45
C4	Network size	0.94	0.10	–0.01	–0.10

Figure 3 Consumer buying behaviour prediction model frame diagram



The index of consumption behaviour extracted above covers a wide range. Therefore, based on Table 2, the parameters corresponding to each prediction index are calculated according to rough set theory (Qiu et al., 2020). The calculation formula is as follows:

$$\hat{t}_{ij}^{*} = \frac{r_{ij}}{\sqrt{\hat{\lambda}_j^{*}}} \tag{9}$$

where $\hat{t}_{ij}^{*}$ represents the j^{th} eigenvalue standard vector of the original variable matrix; r_{ij} is the original variable matrix; λ_j^{*} denotes the j^{th} eigenvalue (Ijabadeniyi and Govender, 2019).

The index weight is calculated according to the ratio of variance contribution rate to cumulative contribution (Xu et al., 2020), and the expression of prediction impact indicators is shown as follows:

$$F = 0.3459 \times F1 + 0.2578 \times F2 + 0.1042 \times F3 + 0.0851 \times F4 \tag{10}$$

The prediction model framework is built on the basis of the prediction influence index expression obtained above. The framework of consumer purchasing behaviour prediction model is shown in Figure 3.

In this study, the model parameters were accurately determined to achieve efficient prediction of consumer buying behaviour. The parameter determination formula is:

$$\begin{cases} hiddennum = \dfrac{U^{T-1}}{F \cdot \hat{t}_{ij}^{*}} \cdot \sqrt{in\dim + out\dim} + n \\ D = (in\dim + 1) \cdot hiddennum + (hiddennum + 1) \cdot out\dim \end{cases} \tag{11}$$

where *hiddennum* represents the consistency parameter; *indim* stands for attribute determination of input data; *outdim* represents the output data attribute determination; *n* denotes auxiliary parameters in the range.

The consumer behaviour is mapped to the low-dimensional manifold (Lee and Lim, 2020), and the consumer behaviour in the low-dimensional prevalence is assumed to be W_L, then the Euclidian coordinates are $(x_1, x_2, \cdots, x_m)$. m represents the dimension of a low-dimensional manifold, and there are k nearest adjacent points near W_L, and H_i represents an arbitrary nearest adjacent point. Then the Euclide coordinate is $(x_1^i, x_2^i, \cdots, x_m^i)$, and l represents the nearest Euclide distance between two points, and its expression formula is:

$$l = \left(\sum_{j-1}^{m} \left(x_j - x_j^i \right)^2 \right)^{\frac{1}{2}} \tag{12}$$

In the formula, the smaller the value of l, the more similar W_L is to the nearest adjacent point. The nearest adjacent point with the most similar value is marked as $H_i^{\min}$. The prediction of consumer behaviour can be realised by the corresponding action of W_L according to $H_i^{\min}$.

4 Experimental study

4.1 Experimental scheme

The following experimental process is designed to verify the effectiveness of the rough set-based prediction method of consumer purchase behaviour in online marketing system.

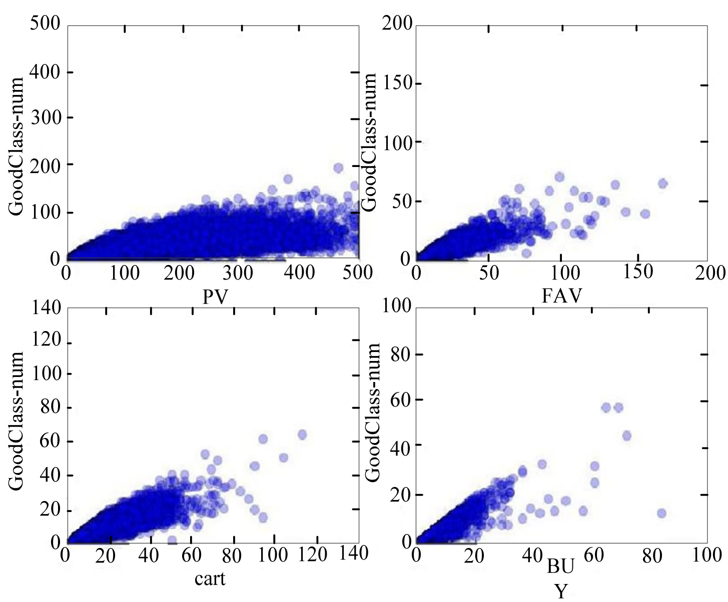
The experiment was completed on the Matlab platform, and the experimental data came from the official data set of Alibaba Lab, including the data between December 1, 2020 and December 31, 2020. The original data set contains 987,568 consumers, with 4,253,561 items of goods and 9,653 categories of goods. The consumption behaviours involved include Page View (PV), BUY (Buy), add to CART (CART) and Favourite (FAV). An example of the raw data is shown in Table 3.

Table 3 Sample raw data

userID	*GoodID*	*GoodclassID*	*Search*	*Time*
1	1531236	2865935	PV	2020-12-5, 09:22:25
1	3862536	4523658	PV	2020-12-10, 15:04:55
1	4425631	2635751	PV	2020-12-24, 21:22:02
1	4652356	2536554	PV	2020-12-26, 15:22:26

The behaviour of consumers in the original data set and the scatter graph of the category of the product in which the behaviour is carried out are analysed. The relationship between consumer behaviour and commodity categories is shown in Figure 4.

Figure 4 Consumer behaviour is associated with commodity categories (see online version for colours)



In the short term, consumers will make repeated purchases of the same kind of goods, which proves that consumers have a high degree of interest in this kind of goods. Consumers will like most of the same kind of goods, and they will not add to the shopping cart. As shown in Figure 4, most of the behaviours generated before purchase are click operations, which involve half of the number of commodity categories. In the four kinds of consumption behaviours, the number of consumer behaviours is always higher than the number of commodity categories generated by the behaviour. Before the final purchase, consumers will conduct several repeated operations, including clicking, liking and adding to the shopping cart, which is the most suitable for the purchase operation, and clicking occupies a high proportion.

4.2 Research on performance indicators

1 Accuracy rate and recall rate of consumer purchasing behaviour demand acquisition. Accuracy and recall are basic indexes. The higher the accuracy and recall are, the higher the effectiveness of the prediction method will be.

Accuracy ratio P:

$$P = \frac{A}{A + B} \tag{13}$$

Recall rate R:

$$R = \frac{A}{A + C} \tag{14}$$

where A is positive class retrieval; B is negative class retrieval; C is the negative class in the positive class retrieval.

2 Deviation of prediction results. The smaller the deviation of the prediction result, the better the prediction effect.

Assuming that the data scale of consumers' purchasing behaviour is N, the true value is *observed*, and the predicted result is *predicted*. The deviation of the forecast results shall be calculated through the following process:

Firstly, the mean square error is calculated using the mean of the sum of squares of the numerical differences between the real value and the predicted result. The calculation formula is as follows:

$$MSE = \frac{1}{N} \sum_{t=1}^{N} \left(predicted_t, observed_t \right)^2 \tag{15}$$

The calculation method of root mean square error is as follows:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(predicted_t, observed_t \right)^2} \tag{16}$$

After analysing the deviation of the predicted values, the deviation of the predicted results can be obtained as follows:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|predicted_t - observed_t\right| \tag{17}$$

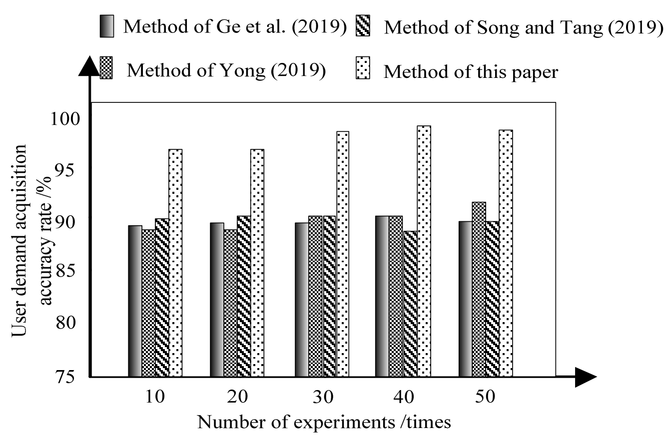
4.3 Results and discussion

In order to avoid too single experimental results, method of Ge et al. (2019), method of Song and Tang (2019) and method of Yong (2019) were taken as comparisons to complete performance verification together with Method of this paper.

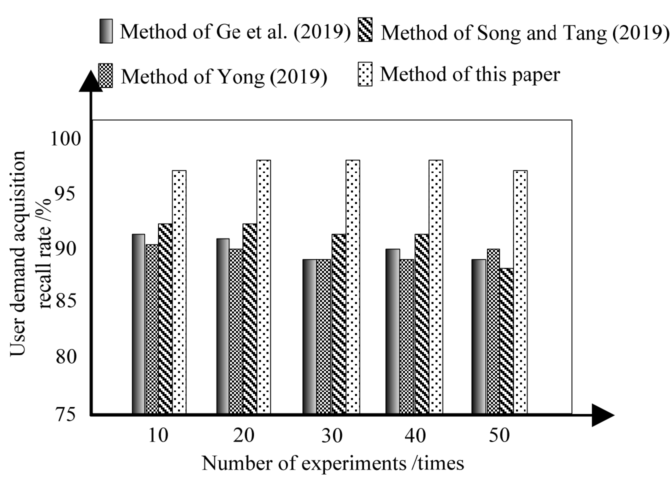
4.3.1 Accuracy and recall test

1 Figure 5 shows the verification results of accuracy and recall of consumer behaviour information by different methods.

Figure 5 Comparison of precision and recall of different methods, (a) accuracy ratio (b) recall rate



(a)



(b)

As shown in Figure 5, the accuracy and recall rates of Method of this paper on consumer behaviour information are higher than those of the three traditional methods, and both of them are higher than 95%, proving that Method of this paper has a good performance. The main reasons for this results are as follows: Method of this paper uses rough set theory to collect data related to consumers' purchasing behaviour, with high collection accuracy, which greatly improves the accuracy and completeness of mining users' implicit demands.

4.3.2 Prediction result deviation test

First of all, RMSE changes were determined in the case of changing window size, and the results were presented in Figure 6.

Figure 6 RMSE change curve diagram

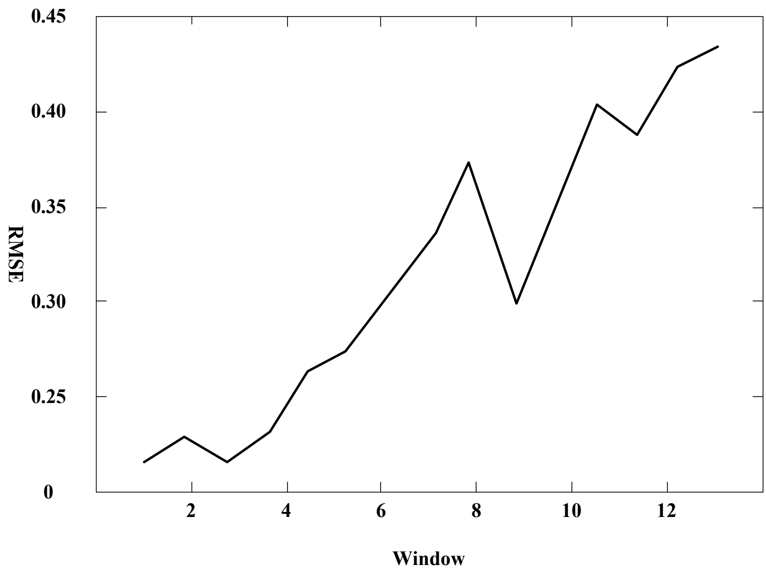


Table 4 Comparison of prediction results deviations by different methods (%)

Number of experiments/times	*Method of this paper*	*Method of reference [5]*	*Method of reference [6]*	*Method of reference [7]*
10	10.23	20.13	19.23	31.02
20	9.56	23.10	12.03	33.01
30	8.12	21.00	16.46	31.46
40	9.12	26.54	12.97	34.16
50	10.03	24.10	23.16	35.05
60	11.03	29.56	22.10	31.00
70	11.00	23.16	23.46	32.00
80	9.45	20.12	25.19	33.45
90	8.16	29.45	28.48	35.79
100	9.15	23.46	27.46	39.12

According to the variation of RMSE in Figure 6, when the window is 10, RMSE is small, indicating the predicted result and the actual result in this case, as shown in Table 4. See Table 4 for details.

As shown in Table 4, the deviation of prediction results of Method of this paper is far lower than that of the three traditional methods, and its minimum value is only 8.12%. The main reasons for this result are as follows: Method of this paper introduces the rough set theory on the basis of the influence index of consumer buying behaviour prediction, which has a unique advantage in dealing with the uncertain and inconsistent information of consumer buying behaviour. By reducing consumer data, it has a strong ability to solve problems, and improves the prediction accuracy by determining the consistency parameters.

5 Conclusions

To solve the large deviation of consumer purchase behaviour prediction and low acquisition accuracy in the traditional online marketing system, a prediction method of consumer purchase behaviour in online marketing system based on rough set is proposed. The innovation of this method is that it reduces the data of consumer purchase behaviour through the control of rough set, reduces the deviation of prediction results by removing the redundant characteristics of network marketing system, and thus fundamentally improves the accuracy and recall rate of prediction method. The experimental results show that the accuracy and recall of this method are higher than 95%, and the minimum deviation of the prediction results can reach 8.12%, which verifies the relative reliability of the prediction results. The era of big data has a great impact on consumers, and more changes will take place in purchase behaviour. The research results of this paper have limited guidance for the prediction of purchase behaviour. The change of purchase behaviour data in the short term is unlikely, so it is necessary to fully consider other variables and fully analyse consumers in the long-term and regulatory changes.

References

Das, D.M. and Ramalingam, D.M. (2019) 'How effective are your actions? Impact of perceived consumer effectiveness on green purchasing behavior', *Academy of Marketing Studies Journal*, Vol. 23, No. 4, pp.52–66.

Davydenko, M. and Peetz, J. (2020) 'Shopping less with shopping lists: planning individual expenses ahead of time affects purchasing behavior when online grocery shopping', *Journal of Consumer Behaviour*, Vol. 19, No. 3, pp.240–251.

Dong, Y. and Jiang, W. (2019) 'Brand purchase prediction based on time-evolving user behaviors in e-commerce', *Concurrency and Computation: Practice and Experience*, Vol. 31, No. 1, pp.1–15.

Ge, S.L., Ye, J. and He, M.X. (2019) 'Prediction model of user purchase behavior based on deep forest', *Computer Science*, Vol. 46, No. 9, pp.190–194.

Gillani, A., Kutaula, S., Leonidou, L.C. and Paul, C. (2019) 'The impact of proximity on consumer fair trade engagement and purchasing behavior: the moderating role of empathic concern and hypocrisy', *Journal of Business Ethics*, Vol. 17, No. 04, pp.1–21.

Ijabadeniyi, A. and Govender, J.P. (2019) 'Coerced CSR: lessons from consumer values and purchasing behavior', *Corporate Communications An International Journal*, Vol. 24, No. 3, pp.515–531.

Lee, C. and Lim, S.Y. (2020) 'Impact of environmental concern on image of internal GSCM practices and consumer purchasing behavior', *Journal of Asian Finance Economics and Business*, Vol. 7, No. 6, pp.241–254.

Lee, H., Kyu, J. and Jin, H. (2019) 'A study on costume jewelry design attributes and consumer purchasing behavior of millennial generation', *Journal of the Korean Society Design Culture*, Vol. 25, No. 1, pp.367–379.

Liu, Y.Y., Li, S. and Li, X.T. (2020) 'Online shopping perceived risk measurement and consumer behavior analysis based on e-commerce data', *Statistics and Management*, Vol. 35, No. 11, pp.77–84.

Qiu, Q., Wang, Y., Qiao, S., Liu, R., Bian, Z.Y., Yao, T. and Nguyen, T.S. (2020) 'Does air pollution affect consumer online purchasing behavior? The effect of environmental psychology and evidence from China', *Journal of Cleaner Production*, Vol. 260, No. 1, pp.16–23.

Song, Z.J. and Tang, X.L. (2019) 'Research on the influence of price and evaluation on consumers' online purchase intention based on S-O-R model', *Practice and Understanding of Mathematics*, Vol. 25, No. 14, pp.1–9.

Toussaint, M., Cabanelas, P. and Alvarado, T. (2020) 'What about the consumer choice? The influence of social sustainability on consumer's purchasing behavior in the food value chain', *European Research on Management and Business Economics*, Vol. 27, No. 1, pp.134–145.

Xu, X., Zeng, S. and He, Y. (2020) 'The impact of information disclosure on consumer purchase behavior on sharing economy platform Airbnb', *International Journal of Production Economics*, Vol. 23, No. 1, pp.107–118.

Yong, S.S. (2019) 'Markov prediction of the impact of commodity promotion on consumers' purchasing intention', *Statistics and Decision*, Vol. 35, No. 9, pp.78–80.