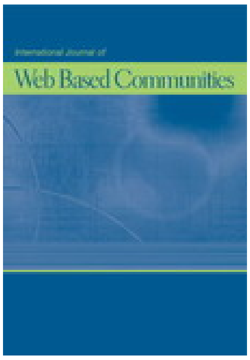


International Journal of Web Based Communities

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

A deep mining method for consumer behaviour data of e-commerce users based on clustering and deep learning

Jing Li

DOI: 10.1504/IJWBC.2023.10048320

Article History:

Received:	04 June 2021
Accepted:	25 November 2021
Published online:	20 January 2023

A deep mining method for consumer behaviour data of e-commerce users based on clustering and deep learning

Jing Li

Puyang Vocational and Technical College,
Puyang 457000, Henan, China
Email: jingli@36haojie.com

Abstract: The data mining accuracy of e-commerce users' consumption behaviour is low and the data clustering effect is poor, so a deep mining method of e-commerce users' consumption behaviour data based on clustering and deep learning is proposed. The consumption behaviour data are divided into simple type, deterministic type, habitual row type and preference type through the user's web browsing log, and the features of the consumption behaviour data are extracted. The centroid and class spacing of behaviour characteristic data are obtained according to the actual distance between the behaviour characteristic data points. The behaviour data deep mining model is built based on the small wave neural network and the deep learning algorithm, and the optimal solution of the model is thus obtained by the gradient descent method, so as to realise the deep mining of the consumption behaviour data. The results show that the accuracy of the proposed method is up to 97%.

Keywords: data clustering; deep learning; dimension kernel function; centroid.

Reference to this paper should be made as follows: Li, J. (2023) 'A deep mining method for consumer behaviour data of e-commerce users based on clustering and deep learning', *Int. J. Web Based Communities*, Vol. 19, No. 1, pp.2–14.

Biographical notes: Jing Li Received her Master's degree in Computer sofTware and Theory from PLA University of Information Engineering in 2007. Currently, she is an Associate Professor in Puyang Vocational and Technical College. Her research interests include data mining and cloud computing.

1 Introduction

With the continuous and in-depth development of electronic information technology, e-commerce and other online enterprises suddenly emerged. The competition among various e-commerce enterprises is increasingly fierce (Li, 2020). With the white-hot development of competition, the consumption behaviour of its potential users has also changed greatly. E-commerce users make comparisons among massive e-commerce products and then make targeted choices based on their cost performance. The consumption behaviour of e-commerce users is an important factor affecting the long-term development of e-commerce enterprises (Arain et al., 2020). Therefore, the consumption behaviour of e-commerce users becomes an intangible asset in the competition of e-commerce enterprises. The economic profits of e-commerce enterprises can be effectively improved by understanding the consumption behaviour of e-commerce users and adjusting the development trend of e-commerce enterprises according to their consumption habits. Therefore, the analysis of data about e-commerce users' consumption behaviour is the key for e-commerce enterprises to remain invincible (Porretta et al., 2018). Therefore, researchers in this field have mined the data of users' consumption behaviour and made some achievements.

Wang et al. (2020) proposed a clustering-based method for mining e-commerce users' consumption privacy behaviour. This method first determines the privacy of consumer behaviour data from electricity users, then the user consumption behaviour of web log data preprocessing, with the help of related algorithm to determine the user data about consumer behaviour, the similar degree between then according to the related algorithm in clustering algorithm rules set mining, implementation of user behaviour data mining. This method is highly efficient in data mining, but is no effective in classifying the relevant behaviour data, with the problem of low mining accuracy. Zuo et al. (2019) proposed to design a data mining model of multi-attribute behaviour of network consumers based on the prospect theory. In this model, the current development status of e-commerce is firstly analysed, and the purchasing process of e-commerce users is effectively analysed with the help of the prospect theory and fuzzy multi-attribute decision-making method. On this basis, the value function, weight function and prospect function are improved, and the user behaviour data are used to make decisions through these functions, and the e-commerce network word-of-mouth data are quantified to complete the design of the mining model of e-commerce users' consumption behaviour. This method is good for analysing the consumption process of e-commerce users, but it takes little account of the consumption behaviour data of consumers and has low data mining accuracy. Zhang et al. (2020) proposed the research on the data mining method of user consumption behaviour based on multi-dimensional behaviour analysis. This method converges the browsing data of e-commerce users to form a multi-dimensional data sequence. With the help of unsupervised feature selection, features in the above sequence are extracted. Then, similarity network is constructed based on similar feature data, and user behaviour data are divided and clustered to complete data mining. This method has a good clustering effect on user behaviour data, but is relatively complex in operation process and is low in work efficiency.

In order to solve the shortcomings of the existing methods, this paper proposes a new method for in-depth data mining of e-commerce users' consumption behaviour. The technical route of this method research is as follows:

Step 1 Divide the consumption behaviour data of e-commerce users into simple data, deterministic data, habit data and preference data through users' web browsing logs; map the data of users' consumption behaviour into different dimensional Spaces, and extract the data features of e-commerce users' consumption behaviour with the help of decision function.

Step 2 Determine the behaviour data through association rules algorithm for association rules, and determine the same correlation degree between user behaviour data with the aid of included angle cosine method; On this basis, determine the characteristics of data similarity, obtain behavioural characteristics data between the centroid and class spacing according to the behaviour characteristics of the actual distance between data points, complete electricity user consumption behaviour characteristics of the data clustering.

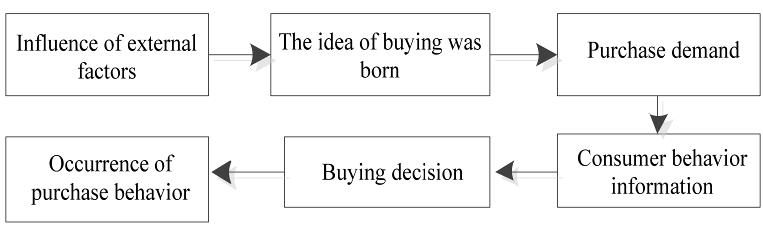
Step 3 Build a deep mining model of e-commerce users' consumption behaviour data through the medium and small wave neural network of deep learning algorithm, and use gradient descent method to solve the optimal solution of consumption behaviour data mining after mining, so as to realise the deep mining of e-commerce users' consumption behaviour data.

2 Pre-processing of consumer behaviour data of e-commerce users

2.1 Feature extraction of consumer behaviour data

E-commerce consumer behaviour mainly refers to the consumer who often consumes on the network. This type of behaviour that the user completes the purchase behaviour through the analysis of e-commerce products, acquisition and decision-making is consumer behaviour. The consumer behaviour of e-commerce users includes: purchased goods, purchase purpose, purchase address and structure decision (Yang, 2020). Every decision-making behaviour of e-commerce users has different nodes. The decision-making process and purchase behaviour of e-commerce users' consumption behaviour are shown in Figure 1.

Figure 1 Consumption of e-commerce users



The process of the formation of e-commerce user consumption behaviour is analysed in detail in Figure 1. By analysing the occurrence of user purchase behaviour, we master the user's consumption behaviour. It plays an important role in guiding the sustainable development of e-commerce enterprises to determine their potential needs according to the process of decision-making and purchase. Therefore, before mining the consumer behaviour data of e-commerce users, the data is collected effectively.

Consumer behaviour data of e-commerce users generally exist in the form of internet logs. As long as e-commerce users log in to relevant shopping websites, there will be behaviour traces, which mainly exist in the form of logs. According to the user's web browsing log, the user's consumption behaviour is divided into simple type, certain type, habitual type and preference type. Therefore, the extraction of consumer behaviour data of e-commerce users is mainly realised through the collection of log data (Hu et al., 2020).

2.2 Pretreatment of consumer behaviour data characteristics

In the data acquisition of consumer behaviour log, kernel function is used to collect user behaviour data. The consumer behaviour data of e-commerce users are mapped by kernel function and reflected in the space with higher dimension, so that the user consumption behaviour log data can be divided linearly in space. On this basis, different dimensions of e-commerce user consumption behaviour data are obtained.

H was set to represent a subset of P^n consumer behaviour dataset, then the mapping of e-commerce user consumption behaviour data H to a fixed space can be expressed as:

$$^{\circ}F = \begin{cases} \{H \in P^n \to H \in Y \\ h \to {}^{\circ}F(h) \end{cases} \tag{1}$$

where $^{\circ}F$ represents the result value of consumer behaviour data mapping of e-commerce users, and Y represents fixed space.

On this basis, the set of behaviour data after mapping the consumption behaviour of e-commerce users is set as follows:

$$R = \{r_1, r_2, ..., r_m\} \tag{2}$$

where r_m represents the composition of consumer behaviour data of e-commerce users.

According to the above determined dataset of e-commerce user consumption behaviour, with the help of the decision function in the kernel function, the characteristics of e-commerce user consumption behaviour data are as follows:

$$z(x) = h \in R^n \to H \tag{3}$$

where $z(x)$ represents the final collection of consumer behaviour data for e-commerce users.

In the data collection of consumer behaviour of e-commerce users, this paper first analyses the essence of consumer behaviour data, maps the consumer behaviour data to different dimensional spaces, and uses the decision function in the kernel function to complete the final data collection (Qian et al., 2020).

Assuming that the dataset of consumer behaviour of e-commerce users is:

$$H = (h_1, h_2, h_3, ..., h_n) \tag{4}$$

where $h_n \sqsubseteq H(n = 1,2\ldots)$ represents the collection of consumer behaviour items for e-commerce users. At this time, the set of e-commerce user consumption behaviour data association rules is set.

The behaviour data H and H' items in the above dataset are inversely proportional to all the behaviour dataset items, and set to:

$$\sup(H \to H') = \frac{(H : h \cup h' \sqsubseteq H' \sqsubset G)}{|G|} \tag{5}$$

The association rules that contain only H in the behaviour data H and H' in formula (5) are determined as follows:

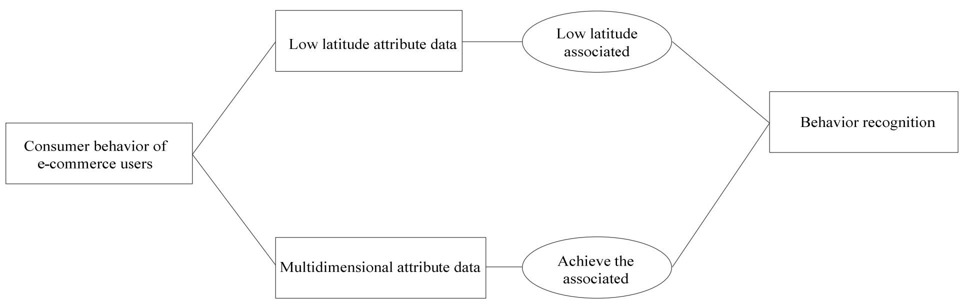
$$con(H \to H') = \frac{H : h \cup h' \sqsubseteq H' \sqsubset G}{|H : H' \sqsubset G|} \tag{6}$$

According to the association rules between the consumption behaviour data, the relationship between the two data in the same set is determined to help remove interference data from the data mining of consumption behaviour (Zhao, 2020), and to determine according to the angle cosine method. That is:

$$\text{cosa} = \frac{a_1 a_2 + b_1 b_2}{\sqrt{a_1^2 + a_2^2}\sqrt{b_1^2 + b_2^2}} \tag{7}$$

Because the consumption behaviour data of e-commerce users change rapidly, there are some limitations in judging by the association rules between their data (Li, 2020; Fang et al., 2020). There are many cross-cutting behaviours in the consumer behaviour, as shown in Figure 2:

Figure 2 There are many kinds of cross behaviours in the consumer behaviour



According to Figure 2, it is necessary to associate multi-dimensional data and determine the differences in consumption behaviours of different e-commerce users. Therefore, deeper mining of the differences in consumption behaviours of different users is needed.

3 Implementation of deep mining of consumer behaviour data of e-commerce users based on clustering and deep learning

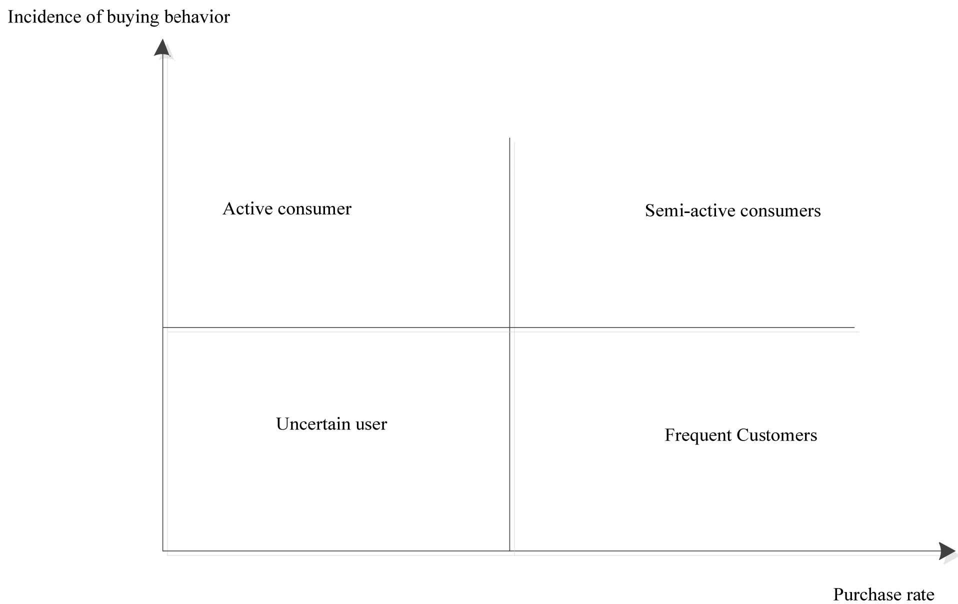
According to the above data, it is difficult to control the consumption behaviour of e-commerce users because the consumption behaviour of users is more changeable. Therefore, in order to realise the deep mining of the consumer behaviour data of e-commerce users, the acquired behaviour characteristics are effectively clustered. Before clustering the characteristics of e-commerce user consumption behaviour data (Yao and Luan, 2019), the user's consumption behaviour is roughly divided into values, as shown in Figure 3.

In the analysis of consumer behaviour data of e-commerce users, the value analysis of users is very important. As long as we master the value concept of consumer groups, we can effectively stimulate the potential consumption behaviour of customers. Therefore, this paper uses clustering algorithm to cluster the data characteristics of e-commerce users' consumption behaviour effectively. Because there are some interference factors in the obtained feature data, it is first processed to reduce the interference degree of feature data clustering. With the help of logical regression method, the processing formula is:

$$c(x) = \frac{1}{1 + Q^{-x}} \tag{8}$$

where Q represents the dependent variable in the data characteristics of e-commerce users' consumption behaviour.

Figure 3 Consumer value types of e-commerce users



Because the dependent variables in the characteristics of consumer behaviour data of e-commerce users will change with the influence of external interference factors, such as commodity price, promotion and other activities, it is necessary to further determine them before carrying out effective clustering is realised by formula (9), that is:

$$q(c(x) = 0) = 1 - \frac{1}{1 + Q^{-kx+d}} \tag{9}$$

where k represents the consumer behaviour of e-commerce users, d represents the bias value.

Based on the above data, the clustering algorithm is used to cluster the consumer behaviour characteristics effectively. Clustering algorithm is an unsupervised method. Its realisation idea is to carry on the multivariate statistics to the research object, and classify by the similarity degree between the sample data. The result of clustering makes each cluster have similar data.

Therefore, in this paper, the similarity between feature data is determined effectively. If the actual distance between the characteristic data points of e-commerce users' consumption behaviour reflects the similarity between the two data, and the characteristic data space S is set as the real vector space of the n dimension S^n, the similarity of e-commerce user consumption behaviour data can be expressed as follows:

$$C_i = (c_i^1, c_i^2, ..c_i^n)^T \tag{10}$$

$$D_i = (d_i^1, d_i^2, ..d_i^n)^T \tag{11}$$

At this point, the distance between the data characteristics of e-commerce users' consumption behaviour in formula (6) and formula (11) can be expressed as follows:

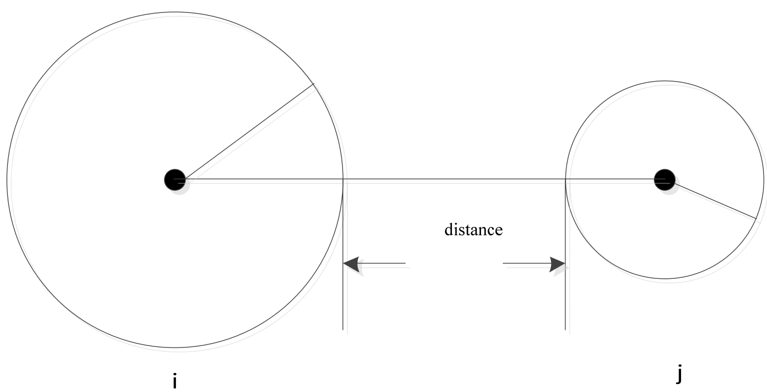
$$Y_L(C_i, D_i) = \left(\sum_{L=1}^{n} | c_i^L - d_i^L |\right)^i \tag{12}$$

After determining the distance between the similar data of the consumer behaviour data of e-commerce users, it is necessary to cluster the similar data with the high degree of similarity according to the determined data centroid, in which, the centroid between the characteristics of consumer behaviour data of e-commerce users is expressed as:

$$v_{i,j} = | v_i - v_j |^2 - c_i - d_i \tag{13}$$

According to the determined centroid, the user consumption behaviour data features can be obtained by class spacing distance, as shown in Figure 4.

Figure 4 Data of user's



On this basis, the data of similar features obtained are divided:

$$s_k = \frac{\sum_{i=1}^{n} V[c^i = l] d_i}{\sum_{i=1}^{n} V[c^i = l]} \tag{14}$$

In the feature clustering of e-commerce user consumption behaviour data, the similarity between e-commerce user consumption behaviour characteristic data points is determined according to the actual distance between e-commerce user consumption behaviour characteristic data points. The centroid and class spacing between e-commerce user consumption behaviour characteristic data are determined to lay the foundation for further mining.

Based on the above clustering of e-commerce users' consumption behaviour characteristics, the same feature data after clustering are analysed in depth. This paper uses deep learning algorithm, including a variety of artificial intelligence algorithms. The deep learning algorithm of wavelet neural network is used to build an e-commerce user consumption behaviour data depth mining model, and the gradient descent method is adopted to obtain the optimal solution of consumer behaviour data mining to realise the deep mining of e-commerce user consumption behaviour data. The algorithm is to determine the relationship between the data in a large number of the same feature dataset, mine the purchase behaviour of e-commerce users, predict the consumption behaviour of users, and thus provide a certain basis for the future sales management of e-commerce enterprises. In this paper, with the help of deep learning algorithm, the wavelet neural network is used to build an e-commerce user consumption behaviour data depth mining model, and the gradient descent method is adopted to obtain the optimal solution of mining consumer behaviour data to achieve e-commerce user consumption behaviour data depth mining.

According to the cross data of consumer behaviour characteristics of e-commerce users obtained above, prediction is made by wavelet neural network (Xia et al., 2020; Wang et al., 2019). According to the different levels of the method, the user consumption behaviour is predicted at a certain time. First, the cross data of consumer behaviour characteristics of e-commerce users are input into the wavelet neural network, and the initial difference values obtained by input layer processing are as follows:

$$U = \sum_{i=1}^{n} R_{j-2k} < U, T_j > + \sum_{k \in N} l - 2k \tag{15}$$

where U and k represents different sequences of cross data input on the consumption behaviour characteristics of e-commerce users, respectively. Based on this, the preliminary prediction value is introduced into the wavelet basis function f for further processing, namely:

$$f = \operatorname{sgn}\left[f_j\right]\left(\left|f_j\right| - t\right) \tag{16}$$

where f represents the time value of the iteration of the wavelet basis function.

Finally, the deep mining model of consumer behaviour data of e-commerce users is constructed, on which the differences of consumer behaviour data of e-commerce users are excavated, that is:

$$W(i) = Q_i \left(\frac{\sum_{i=1}^{n} bf}{f} \right) \tag{17}$$

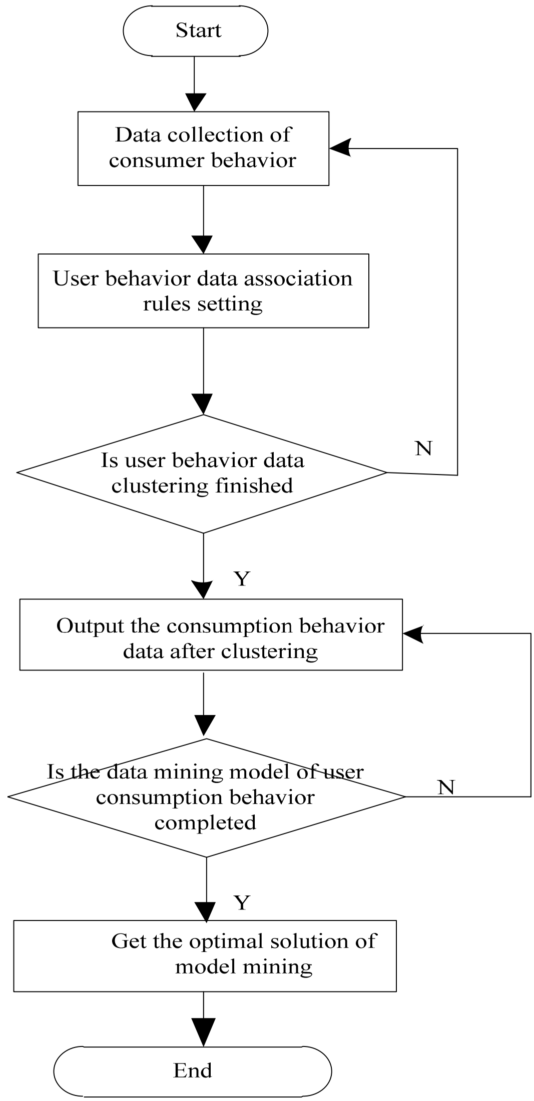
where $H(i)$ represents the final output value of the excavated differential feature data.

In order to further improve the accuracy of mining, this paper uses the gradient descent method to solve the optimal solution of the deep mining model of consumer behaviour data of e-commerce users. The results are as follows:

$$q(x) = \frac{e(k) - \gamma}{p} \tag{18}$$

where $e(k)$ represents the consumer behaviour data of the original ecommerce user, γ represents the mean of the original value and p standards difference representing the characteristics of differential data. The data mining process of consumer behaviour of e-commerce users is shown in Figure 5.

Figure 5 Data mining for consumer behaviour



According to the above clustering user consumption behaviour data, the depth mining model of e-commerce user consumption behaviour data is constructed by the wavelet neural network of depth learning algorithm, and the optimal solution of mining consumption behaviour data after mining is obtained by gradient descent method. The depth mining of consumer behaviour data is realised.

4 Experimental analysis

4.1 Experimental scheme

According to the above design method, in order to verify the effectiveness of the method, it is analysed experimentally. The experiment selected 40 employees, including 15 male employees and 35 female employees. The data were collected and placed on the MATLAB platform. The system of data processing and analysis in the experiment was Windows 10 system, and the memory was 16 GB, with SPSS software to obtain data. The detailed parameter design in the experiment is shown in Table 1:

Table 1 Detailed experimental parameters

Parameter	*Data*
Online shopping frequency/week	10
Login time	7:00–8:00,17:00以后
Page length/min	> = 5
Purchase rate/%	> 60
Sample data/GB	2
Number of times of behavioural data mining	100

4.2 Experimental indicators

On the basis of the above experimental environment and parameter design, the experiment is carried out in a comparative way, and the methods of this paper, Zuo et al. (2019) and Zhang et al. (2020) are analysed, and the accuracy of mining and the error of data clustering are taken as the experimental indexes. Among them, the precision calculation formula of mining is:

$$Z = \frac{Z_i}{z_{all}} \times 100\% \tag{19}$$

where Z_i represents the amount of consumer behaviour data mining, z_{all} represents the total amount of consumer behaviour data mining for e-commerce users.

4.3 Analysis of experimental results

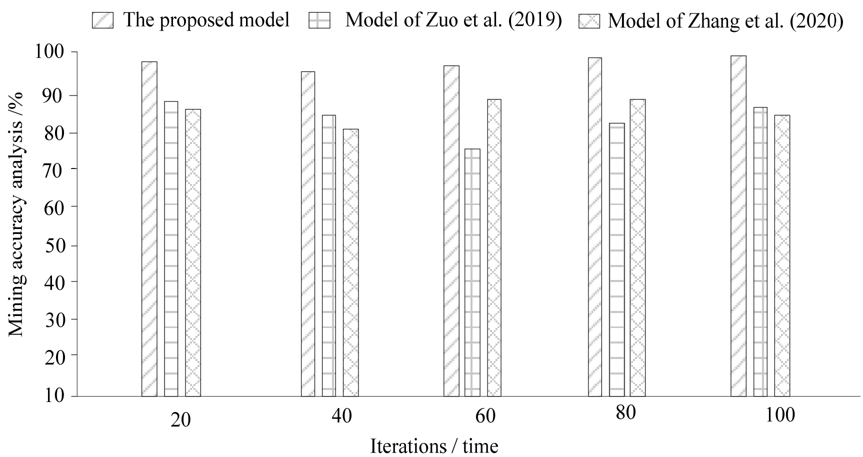
4.3.1 Precision analysis of consumer behaviour data of e-commerce users

The accuracy of deep mining of consumer behaviour data is a direct index to reflect the effectiveness of mining algorithm. To this end, the experimental analysis of this method, Zuo et al. (2019) method and Zhang et al. (2020) method are adopted to sample e-commerce user consumption behaviour data, and the mining accuracy results are shown in Figure 6:

Analysis of the experimental results in Figure 6 shows that as the number of iterations changes, there are some differences in the accuracy of data mining of sample e-commerce users' consumption behaviour among the method proposed in this paper and methods of Zuo et al. (2019) and Zhang et al. (2020). Among them, When the number of iterations is

60, the mining accuracy of the proposed method, the Zuo et al. (2019) method, and the Zhang et al. (2020) method are about 95%, 75%, and about 89%, respectively. When the number of iterations is 100, The mining accuracy of the three methods is about 96%, 85% and 83%, respectively. In contrast, the mining accuracy of the proposed method is always more than 90%, higher than that of the other two methods. This is due to the proposed method can effectively collect the consumer behaviour data of e-commerce users, and can build a mining model with the help of deep learning algorithm, so can deliver a higher efficacy.

Figure 6 Comparison of data depth mining accuracy of e-commerce users' consumption behaviour

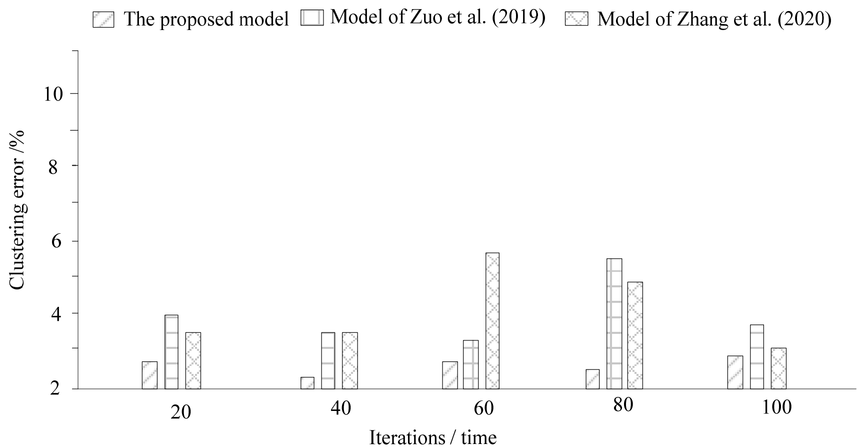


4.3.2 Clustering error of consumer behaviour data of e-commerce users

In the process of deep mining of consumer behaviour data, clustering of consumer behaviour data is an important factor affecting its mining performance. Therefore, the error of this method, Zuo et al. (2019) method and Zhang et al. (2020) method on the data clustering of sample e-commerce users' consumption behaviour is analysed experimentally. The results are shown in Figure 7:

Analysis of the data in Figure 6 shows that as the number of iterations changes, there are some differences in the error of clustering data of consumer behaviour of e-commerce users among the method proposed in this paper and methods of Zuo et al. (2019) and Zhang et al. (2020). When the number of iterations is 40, the error of clustering data of user consumption behaviour of the method proposed in this paper, the Zuo et al. (2019) method and the Zuo et al. (2019) method is 2.2%, about 3.5%, and about 3.5%, respectively; When the number of iterations is 60, the error of clustering data of user consumption behaviour of the three methods is 2.6%, about 3.1%, and about 5.8%, respectively; When the number of iterations is 100, the error of clustering data of user consumption behaviour of the three method is 2.7%, about 3.3%, and about 3.1%, respectively. The lower clustering error of the proposed method verifies its effectiveness.

Figure 7 Clustering error analysis of consumer behaviour data of e-commerce users



5 Conclusions

In order to improve the sales quality of e-commerce users, a deep mining method based on clustering and deep learning is proposed. According to the actual distance between the behaviour feature data points, the similarity of the feature data is determined, the centroid and class spacing between the behaviour feature data are obtained, and the e-commerce user consumption behaviour data feature clustering is completed; Through the deep learning algorithm and wavelet neural network, an e-commerce user consumption behaviour data depth mining model is built, and the gradient descent method is used to solve the optimal solution of mining consumer behaviour data to realise the deep mining of e-commerce user consumption behaviour data. This method has the following advantages

1 The accuracy of the proposed method is about 97%, which has a certain degree of credibility

2 The proposed method has advantageous low clustering error for consumer behaviour data of e-commerce users.

Acknowledgements

This work was supported by Key R&D and Promotion Special Project (Science and Technology) Foundation of Henan Province in 2020 under grant no.212102210081, and Higher Education Teaching Reform Research and Practice Project of Henan Province in 2019 under grant no. 2019SJGLX666.

References

Arain A.L., Pummill, R. and Adu-Brimpong, J. (2020) 'Analysis of e-waste recycling behavior based on survey at a Midwestern US University', *Waste Management*, Vol. 105, No. 14, pp.119–127.

Fang, C., Guo, Y-B., Wang, N., Zhen, S-H. and Tang, G-D. (2020) 'Differential private data publishing method based on generative adversarial network', *Acta Electronica Sinica*, Vol. 48, No. 10, pp.1983–1992.

Hu, J., Zou, M., Wen, W. and Zhou, Z. (2020) 'Research process of big-data-driven learning analysis technology', *Modern Electronics Technique*, Vol. 43, No. 18, pp.54–58.

Li, Y. (2020) 'Data mining algorithm based access control model for users of mobile e-commerce community', *Modern Electronics Technique*, Vol. 43, No. 4, pp.153–156.

Li, Z. (2020) 'Consumer behavior analysis model based on machine learning', *Journal of Intelligent and Fuzzy Systems*, Vol. 14, No. 3, pp.1–11.

Porretta, S., Gere, A. and Radványi, D. (2018) 'Mind genomics (conjoint analysis): the new concept research in the analysis of consumer behaviour and choice', *Trends in Food Science and Technology*, Vol. 14, No. 52, pp.45–49.

Qian, Z., Zhou, T., Tao, F., Fan, Z-L., Liu, H. and Lu, J. (2020) 'Ouster based extraction method of entrances and exits in AOI', *Geography and Geo-Information Science*, Vol. 36, No. 2, pp.1–7.

Wang, J., Chen, Z., Liu, J. and Cheng, H. (2020) Privacy behavior mining technology for cloud computing based on clustering', *Computer Engineering and Applications*, Vol. 56, No. 5, pp.80–84.

Wang, L., Gao, S., Li, J. and Fan, H. (2019) 'The sales character analysis of retails based on K-means method and spatial correlation', *Bulletin of Surveying and Mapping*, Vol. 14, No. 9, pp.51–54.

Xia, T., Shi, Z. and Lai, L. (2020) 'Study on the relationship between the social exclusion and pro-social consumer behavior in the conformity context', *Journal of Management Science*, Vol. 33, No. 1, pp.114–125.

Yang, J. (2020) 'The legal inclusion of fair use to text and data mining', *Library Tribune*, Vol. 44, No. 4, pp.141–150.

Yao, Y. and Luan, W. () 'Factors Influencing yacht tourism consumption behavior based on the TAM-IDT model', *Tourism Tribune*, Vol. 34, No. 2, pp.60–71.

Zhang, L-B., Guo, Q., Wu, X-B., Liang, Y-Z. and Liu, J-G. (2020) 'User clustering method based on multi-dimensional behavior analysis', *Journal of University of Electronic Science and Technology of China*, Vol. 49, No. 2, pp.315–320.

Zhao, B., Han, J., Shi, T., Ji, G., Liu, X. and Yu, Z. (2020) 'Advancements in semantic trajectory modelling and mining', *Journal of Geo-Information Science*, Vol. 22, No. 4, pp.842–856.

Zuo, W-M., Chen, S-J., Wang, X. and Chen, H-Q. (2019) 'Multiple attribute behavioral decision-making model of internet consumers based on prospect theory', *Journal of Industrial Engineering and Engineering Management*, Vol. 33, No. 3, pp.125–135.