International Journal of Web Based Communities

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

Cluster analysis-based big data mining method of e-commerce consumer behaviour

Lei Xue

DOI: 10.1504/IJWBC.2023.10048319

Article History:

Received:	21 May 2021
Accepted:	25 November 2021
Published online:	20 January 2023

Cluster analysis-based big data mining method of e-commerce consumer behaviour

Lei Xue

School of Economics and Management,
Henan Polytechnic Institute,
Nanyang, 473000, China
Email: leixue@36haojie.com

Abstract: In order to overcome the problems of low precision and long time of data mining in traditional big data mining methods of consumer behaviour, a clustering analysis method for big data mining of e-commerce consumer behaviour is proposed. In this paper, the K-means algorithm is used to calculate the similarity of behaviour clustering nodes of fee payers, determine the clustering process of consumer behaviour data and determine the mining weight of behaviour data. According to the FCM clustering algorithm, the target function for data mining of e-commerce consumer behaviour is constructed. According to Lagrange multiplication, the membership degree of consumer behaviour data is obtained, and the big data mining of consumer behaviour in e-commerce is realised. The experimental results show that with the method proposed in this paper, when the number of consumers is 500, the time for big data mining of consumer behaviour is 15.6s and the accuracy of behaviour big data mining is 95.34%.

Keywords: data mining; K-means clustering analysis; FCM clustering algorithm; Lagrange multiplication; cluster node.

Reference to this paper should be made as follows: Xue, L. (2023) 'Cluster analysis-based big data mining method of e-commerce consumer behaviour', *Int. J. Web Based Communities*, Vol. 19, No. 1, pp.53–63.

Biographical notes: Lei Xue received her Master's degree from Henan University, China. Currently, she works in School of Economics and Management, Henan Polytechnic Institute. Her research interests include ideological and political education and Marxist theory.

1 Introduction

The rapid development of network technology has led to the generation of a lot of data, and the emergence of shopping software such as Taobao, Jingdong, and pinduoduo makes online sales more common, and makes the development of e-commerce industry more rapid (Pb and Ip, 2020; Jiang et al., 2019; Khani et al., 2021). In response to this phenomenon, how to analyse the behaviour of e-commerce users through the big data of e-commerce consumer behaviour, obtain the characteristic attributes of the data, and realise the accurate prediction of user consumption tendency? (Doryab et al., 2019) Relevant scholars have made some progress by using various technologies.

Wei and Wang (2019) proposed an e-commerce user behaviour analysis method. By collecting big data of user consumption behaviour on online shopping platform, single factor and multi factor were used to obtain user behaviour probability and user attribute characteristics; user operation mode was used to obtain commodity cost performance attribute characteristics, and K-means method was used to calculate user behaviour probability and consumption demand characteristics, In terms of the analysis of e-commerce user behaviour characteristics, this method can obtain more accurate user characteristics, but the accuracy of user consumption behaviour mining is poor. Liu and Zhang (2019) obtained the user behaviour data of e-commerce platform and analysed the product features of e-commerce users by the data mining method, eliminated the redundant features of user consumption behaviour by data processing technology, made word frequency statistics of e-commerce users' consumption behaviour by text analysis method, obtained and classified the differentiated features of consumption behaviour, This method can help accurately realise the user demand analysis, but it takes a long time to mine the user's consumption behaviour. Li et al. (2019a) proposed a mining and analysis method for poor students' consumption big data. They obtained the consumption data of poor students through big data, used data mining method to collect consumption data features, built a shared data warehouse for poor students' consumption big data, designed consumption feature mining workflow, and realised intelligent decision-making and consumption mining analysis, But the recall rate of this method is not good.

In order to solve the problems of the above methods, this paper proposes a big data mining method of e-commerce consumer behaviour based on cluster analysis. The specific research ideas are as follows:

Firstly, the K-means algorithm is used to determine the similarity of clustering nodes, and the clustering process of e-commerce consumer behaviour data is given;

Secondly, K-means clustering analysis method is used to obtain the attributes of e-commerce consumer behaviour data and determine the weight of consumer behaviour data.

Then, the FCM clustering algorithm is used to construct the objective function of e-commerce consumer behaviour data mining, and the membership degree of consumer behaviour data is obtained by Lagrange multiplication to realise e-commerce consumer behaviour big data mining.

Finally, through the relevant experiments, e-commerce consumer behaviour big data mining experiments are carried out, and conclusions are drawn and summarised in the full text.

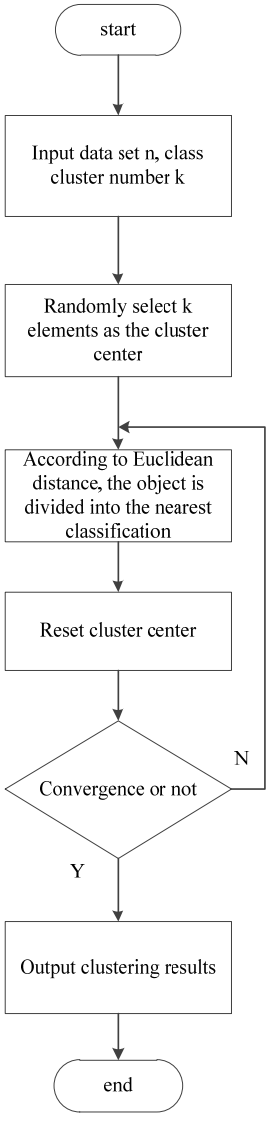
2 Clustering method of consumer behaviour data based on K-means algorithm

2.1 K-means clustering analysis

The k-means algorithm is used to divide the data into k sets according to the given K value, that is, the number of clusters to be divided in advance, which is also called cluster (Sarker, 2019; Li et al., 2019b, 2020; Sun et al., 2020; Desai et al., 2021). The basic process of K-means algorithm is shown in Figure 1. For a data set containing x elements, a constant K should be given first, and the value of K is the number of clusters generated by the final clustering. Among all the elements, k elements should be found randomly

and set as the initial cluster centre. Then, the distance between each element and the pre-selected cluster centre is calculated and compared one by one, and the element is divided into the cluster where the nearest cluster centre is located. After calculating and dividing clusters, the centre of the new cluster is calculated again to find the new cluster centre, and then the above steps are repeated to transform the cluster centre and cluster; When the cluster centre is no longer updated, that is, all clusters are fixed and no longer changed, then the clustering process ends. The clustering process of consumer behaviour data is shown in Figure 1.

Figure 1 Consumer behaviour data clustering process



Analysis of Figure 1 shows that the process of consumer behaviour data mining using k-means algorithm is as follows:

- Input: a collection of D consumer behaviour data elements. The K-means clustering method is used to determine the number of clusters.
- Output: a collection of K cluster consumer behaviour data elements.
  1 Select K elements and use cluster analysis method to select the cluster centre of D data elements.
  2 Calculate the weight of different consumer behaviour data elements, get the consumer behaviour data cluster centre, get the new cluster of consumer behaviour characteristics, and then recalculate the consumer behaviour big data cluster centre.
  3 Repeat.
  4 When the consumer behaviour big data cluster centre is constant, the weight calculation process of behaviour big data element partition is ended.

2.2 Clustering of consumer behaviour big data

By calculating the similarity of consumer behaviour big data features, K-means clustering operation is used to solve the centre distance of consumer behaviour data cluster.

Suppose that the consumer behaviour big data set N contains n elements, namely $N = \{a_1, a_2,\ldots, a_i,\ldots a_n\}$, and each consumer behaviour big data set has n attribute characteristics of consumer behaviour data, $a_i = \{p_{i1}, p_{i2},\ldots, p_{ij},\ldots p_{in}\}$, the data in the above two sets are continuous (Shuvo et al., 2019; Mwansa et al., 2020; Ziaei-Bideh and Namakshenas-Jahromi, 2020; Gokulapriya and Kumar, 2020). In order to obtain the correlation between different consumer behaviour data a_e and a_t, we can calculate the physical distance $Dist(a_e, a_t)$ of consumer behaviour data, and use the physical distance value to judge the similarity of consumer behaviour big data. The large the physical distance of different consumer behaviour data, the less similarity between the consumer behaviour data a_e and a_t. The smaller the $Dist(a_e, a_t)$ value of consumer behaviour data, the greater the similarity between a_e and a_t. In this case, the formula for the physical distance $Dist(a_e, a_t)$ of behaviour data is:

$$\mathrm{Dist}(a_e, a_t) = \sqrt{\sum_{r=1}^{n} (a_{er} - a_{tr})^2} \tag{1}$$

When the similarity of a_e and a_t is large, the Mahatma distance is used to calculate the weight distance of different consumer behaviour data:

$$\mathrm{Dist}(a_e, a_t) = |a_{e1} - a_{t1}| + |a_{e2} - a_{t2}| + \ldots + |a_{en} - a_{tn}| \tag{2}$$

When the similarity of a_e and a_t is small, the Minkowski distance is used to calculate the similarity of clustering nodes of different consumer behaviour data:

$$\mathrm{Dist}(a_e, a_t) = \sqrt[d]{\sum_{r=1}^{n} (a_{er} - a_{tr})^d} \tag{3}$$

At this time, the k-means algorithm is used to calculate the similarity of consumer behaviour clustering nodes to obtain the attributes of consumer behaviour data, so as to determine the weight of behaviour data mining.

3 E-commerce consumer behaviour big data mining based on cluster analysis

In order to realise the accurate mining of big data of e-commerce consumer behaviour and improve the feature mining of consumer behaviour big data, based on Section 2, this paper conducts the clustering selection of consumer behaviour feature data points, and uses FCM clustering method to obtain e-commerce consumer behaviour node clustering degree function (Wu et al., 2020; Qian and Zhou, 2019; Zhao and Xu, 2019).

It is assumed that the data set of e-commerce consumer behaviour big data samples is represented by $X = \{x_1, x_2,\ldots x_n\}$, and data attributes are divided by FCM clustering analysis. When the corresponding e-commerce consumer behaviour is obtained according to the feature type, there are c class centres, and C exist in total. In these consumer behaviour big data samples, μ_{ji} is used to represent the membership of the i data in class j (Sotomayor-Bonilla et al., 2021). At this time, the constraint conditions of the mining function J_m and its mining function of consumer behaviour big data target are as follows:

$$J_m = \sum_{i=1}^{C}\sum_{j=1}^{n}\mu_{ij}^{m}\left|X_j - c_j\right|^2 \tag{4}$$

$$\sum_{i=1}^{C}\mu_{ij} = 1, \forall j = 1,2,..,n \tag{5}$$

In formula (4), c is used to represent the number of consumer behaviour big data feature categories, and $1 < c < n$; X is used to represent the number of samples in the consumer behaviour data, and μ_{ji} is used to represent the membership degree of the j^{th} sample of the consumer behaviour big data to the i^{th} class, and $0 \le \mu_{ji} \le 1$; m is the smoothing index of the objective function of behavioural big data, $m \ge 1$; c_j is the clustering centre of behavioural big data set; $|X_j - c_j|^2$ is used to represent the error measure of consumer behaviour characteristics. In order to reduce the error of consumer behaviour big data feature mining, FCM clustering is used to obtain the square value of error, and Lagrange multiplication is used to cluster the consumer behaviour feature data points.

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ji}\left|X_j - C_i\right|^2 + \lambda_1\left(\sum_{i=1}^{C}\mu_{ij} - 1\right) + \ldots + \lambda_n\left(\sum_{i=1}^{C}\mu_{in} - 1\right) \tag{6}$$

The membership degree of consumer behaviour characteristics is obtained by data derivation, and the weight value is:

$$\mu_{ij} = \sum_{k=1}^{C}\left(\frac{\left|X_j - c_i\right|}{\left|X_j - C_k\right|}\right)^{-\left(\frac{2}{m-1}\right)} \tag{7}$$

At the same time, we use the function derivation to obtain the consumer behaviour big data mining function, and get the formula (8):

$$c_i = \frac{\sum_{j=1}^{n} \left(x_j \mu_{ij}^m \right)}{\sum_{j=1}^{n} \mu_{ij}^m} \tag{8}$$

The analysis of formula (7) and formula (8) shows that when the smoothing index of consumer behaviour target mining function is $m \rightarrow 1$, the fuzzy effect of consumer behaviour feature mining function will gradually weaken with FCM clustering analysis method; When the smooth index m of the consumer behaviour target mining function is $m = 1$, the consumer behaviour feature mining function becomes K-means clustering analysis algorithm. Therefore, it is necessary to constrain the value of the smoothing index m of the objective mining function, and under this constraint, the big data mining function of consumer behaviour is obtained, and the final big data mining result of consumer behaviour is obtained.

4 Experiment

4.1 Experimental scheme design

This paper uses MATLAB software to process the big data of e-commerce users' consumption behaviour, and uses Windows 7 as the operating system. The e-commerce consumption data used in the experiment comes from the user data of an app. In order to improve the effectiveness and accuracy of the experimental results, it is necessary to remove the redundant consumption characteristics of users, realise data feature filtering and empty data removal. This paper uses the methods of Wei and Wang (2019), Liu and Zhang (2019), Li et al. (2019a) and this paper to analyse 50 groups of consumer behaviour data of e-commerce users, separately and carry out experimental verification.

4.2 Experimental index

1 Data mining time of consumer behaviour in E-commerce.

   E-commerce consumer behaviour data mining time can reflect the efficiency of e-commerce consumer behaviour data mining. The shorter the time of e-commerce consumer behaviour data mining, the higher the efficiency of e-commerce consumer behaviour data mining. On the contrary, the longer the consumer behaviour data mining time, the lower the efficiency of e-commerce consumer behaviour data mining.

2 Accuracy of big data mining on consumer behaviour.

   The accuracy of behaviour big data mining reflects the effect of different methods on e-commerce consumer behaviour data mining. A higher accuracy of behaviour big data mining indicates better effect of behaviour data mining. On the contrary, a lower accuracy of behaviour big data mining indicates worse effect of behaviour data mining.

3 Consumer behaviour big data mining recall.

The higher the recall rate of consumer behaviour big data mining, the better the effect of consumer behaviour big data mining. On the contrary, the higher the recall rate of consumer behaviour big data mining, the worse the effect of consumer behaviour big data mining.

4.3 Result analysis

4.3.1 Data mining time

In order to verify the efficiency of e-commerce consumer behaviour big data mining under different methods, this paper detects the mining time of the methods of Wei and Wang (2019), Liu and Zhang (2019), Li et al. (2019a) and this paper in mining e-commerce consumer behaviour big data. The results of big data mining of e-commerce consumer behaviour with different methods are shown in Table 1.

Table 1 E-commerce consumer behaviour big data mining time

Number of consumers/person	*E-commerce consumer behaviour big data mining time/s*			
	Wei and Wang (2019) method	*Liu and Zhang (2019) method*	*Li et al. (2019a) method*	*The method*
100	56.5	54.8	66.2	6.8
200	89.65	99.7	96.4	8.8
300	167.5	173.2	165.7	10.3
400	187.6	197.3	187.2	12.6
500	226.8	215.2	267.3	15.6
600	266.2	287.4	299.3	17.0
700	316.7	322.5	352.9	19.5
800	387.4	369.3	401.6	20.8
Mean value	212.3	214.9	229.6	13.9

According to the analysis of Table 1, the mining time of e-commerce consumer behaviour big data is different under different methods. When the number of consumers is 100, the time of e-commerce consumer behaviour big data mining of Wei and Wang (2019) method is 56.5s, the time of e-commerce consumer behaviour big data mining of literature [6] method is 54.8s, the time of e-commerce consumer behaviour big data mining of Li et al. (2019a) method is 66.2s, and the time of e-commerce consumer behaviour big data mining of this method is 6.8s. When the number of consumers is 500, the time of e-commerce consumer behaviour big data mining of Wei and Wang (2019) method is 226.8s, the time of e-commerce consumer behaviour big data mining of Liu and Zhang (2019) method is 215.2s, the time of e-commerce consumer behaviour big data mining of Li et al. (2019a) method is 267.3s, and the time of e-commerce consumer behaviour big data mining of this method is 15.6s. The time of e-commerce consumer behaviour big data mining of this method is far less than other methods, which shows that the efficiency of this method is higher.

4.3.2 Behavioural big data mining accuracy

In order to verify the accuracy of e-commerce consumer behaviour big data mining under different methods, the methods of Wei and Wang (2019), Liu and Zhang (2019), Li et al. (2019a) and this paper are used to test the accuracy of e-commerce consumer behaviour big data mining. The comparison results of the four methods are shown in Table 2.

Table 2 Precision of behavioural big data mining

Number of consumers/person	*Precision of big data mining for consumer behaviour in e-commerce/%*			
	Wei and Wang (2019) method	*Liu and Zhang (2019) method*	*Li et al. (2019a) method*	*The method*
100	82.32	87.87	85.43	99.80
200	80.54	85.90	83.00	98.25
300	78.77	82.42	81.65	98.06
400	75.66	79.32	75.43	96.93
500	72.52	72.32	73.21	95.34
600	68.54	70.26	70.66	94.65
700	62.22	65.66	68.96	93.92
800	58.39	62.44	76.43	93.28
Mean value	72.37	75.77	76.85	96.28

According to Table 2, the accuracy of e-commerce consumer behaviour big data mining is found in different methods. When the number of consumers is 300, 78.77% of the accuracy of e-commerce consumer behaviour big data mining in the method of Wei and Wang (2019), 82.42% of the accuracy of e-commerce consumer behaviour big data mining in Liu and Zhang (2019), 81.65% of the accuracy of e-commerce consumer behaviour big data mining in document (Li et al., 2019a). The accuracy of the method is 98.06% of the accuracy of big data mining of e-commerce consumer behaviour. When the number of consumers is 800, the accuracy of e-commerce consumer behaviour big data mining in Wei and Wang (2019) method is 58.39%, that of the method of Liu and Zhang (2019) is 62.44%, and 76.43% of the accuracy of e-commerce consumer behaviour big data mining in document (Li et al., 2019a). The accuracy of the method is 93.28%. The accuracy of big data mining of consumer behaviour in this method is significantly higher than that of other methods. It shows that the accuracy of behaviour big data mining is high.

4.3.3 Behavioural big data mining recall rate

In order to verify the recall rate of behaviour big data mining under different methods, the methods of Wei and Wang (2019), Liu and Zhang (2019), Li et al. (2019a) and this paper are used to test the recall rate of e-commerce consumer behaviour big data mining. The results of different methods are shown in Table 3.

Table 3 E-commerce consumer behaviour big data mining recall rate

Consumer behaviour characteristic data volume/GB	*E-commerce consumer behaviour big data mining recall rate/%*			
	Wei and Wang (2019) method	*Liu and Zhang (2019) method*	*Li et al. (2019a) method*	*The method*
100	67.8	87.2	87.4	98.6
200	78.4	76.3	88.2	99.5
300	67.3	69.5	83.6	97.3
400	78.3	76.3	83.2	99.4
500	76.3	76.3	88.5	97.5
600	62.1	66.7	73.2	96.3
700	54.3	68.5	55.3	95.4
800	58.5	74.3	65.7	92.6
Mean value	67.9	74.4	78.1	97.2

Analysis of Table 3 shows that the recall rate of e-commerce consumer behaviour big data mining is different under different methods. When the amount of consumer behaviour characteristic data is 200GB, the recall rate of e-commerce consumer behaviour big data mining based on Wei and Wang (2019), Liu and Zhang (2019), Li et al. (2019a) is 78.4%, 76.3%, and 88.2%, respectively, and the recall rate of the method proposed in this paper is 99.5%. When the amount of consumer behaviour characteristic data is 800GB, the recall rate of e-commerce consumer behaviour big data mining based on Wei and Wang (2019), Liu and Zhang (2019), Li et al. (2019a) is 58.5%, 74.3%, and 65.7%, respectively, and the recall rate of the method proposed in this paper is 92.6%. This method proposed in this paper always has a high recall rate of e-commerce consumer behaviour big data mining, which shows that the data mining ability of this method is strong.

5 Conclusions

This paper proposes an e-commerce consumer behaviour big data mining method based on cluster analysis, which uses k-means cluster analysis to obtain consumer behaviour data attributes, and constructs behaviour data mining objective functions according to FCM clustering algorithm; according to Lagrange, the daily multiplication method is used to obtain the membership degree of consumer behaviour data, and realise the big data mining of e-commerce consumer behaviour. The following conclusions are drawn through experiments:

1 When the number of consumers is 500, the e-commerce consumer behaviour big data mining method in this paper takes 15.6s, suggesting that the consumer behaviour big data mining method proposed in this paper is more efficient.

2 When the number of consumers is 800, the accuracy of big data mining of e-commerce consumer behaviour proposed in this paper is 93.28%, suggesting that the behavioural big data mining method of this paper has high accuracy.

3 When the amount of consumer behaviour characteristic data is 800GB, the recall rate of the large data mining of e-commerce consumer behaviour proposed in this paper is 92.6%, suggesting that the data mining ability of the method proposed in this paper is strong.

References

Desai, V.P., Oza, D. and Naik, D. (2021) 'Data mining approach for cyber security', *International Journal of Computer Applications Technology and Research*, Vol. 10, No. 1, pp.35–41.

Doryab, A., Villalba, D.K., Chikersal, P., Dutcher, J., Tumminia, M., Liu, X., Cohen, S., Creswell, K., Mankoff, J. and Creswell, J. (2019) 'Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data', *JMIR mHealth and uHealth*, Vol. 7, No. 7, pp.12–28.

Gokulapriya, R. and Kumar, R.G. (2020) 'Normal discriminant deep convolution neural classification based web behavioral pattern mining for user identification', *Journal of Ambient Intelligence and Humanized Computing*, Vol. 21, No. 05, pp.66–72.

Jiang, L., Huang, G., Huang, C. and Wang, W. (2019) 'Data mining and optimization of a port vessel behavior behavioral model under the internet of things', *IEEE Access*, Vol. 12, No. 04, pp.4–16.

Khani, S., Haghighat, F., Panchabikesan, K. and Ashouri, M. (2021) 'Extracting energy-related knowledge from mining occupants' behavioral data in residential buildings', *Journal of Building Engineering*, Vol. 39, No. 6, pp.102–119.

Li, A., Liang, B. and Li, X. (2019a) 'Mining and analysis of poor students based on big data of student all-in-one card consumption', *Computing Technology and Automation*, Vol. 38, No. 02, pp.130–134.

Li, X., Tong, B.L. and Wu, X.B. (2019b) 'Research on e-commerce customer data mining based on fuzzy operation', *Journal of Sichuan University of Arts and Science*, Vol. 29, No. 05, pp.21–25.

Li, X., Fei, Y., Rizzuto, T.E. and Yang, F. (2020) 'What are the occupational hazards of construction project managers: a data mining analysis in China', *Safety Science*, Vol. 134, No. 2, pp.55–67.

Liu, N. and Zhang, Q. (2019) 'Analysis of consumer demand and product data mining technology based on e-commerce platform', *Inner Mongolia Statistics*, Vol. 18, No. 01, pp.38–41.

Mwansa, M., Hurst, W. and Shen, Y. (2020) 'A framework for encouraging consumer behavioral change by modeling smart meter data energy usage', *International Journal of Big Data Mining for Global Warming*, Vol. 27, No. 16, pp.205–216.

Pb, A. and Ip, A. (2020) 'Data mining approach in evaluation of sustainable entrepreneurship', *Procedia Computer Science*, Vol. 176, No. 21, pp.2725–2735.

Qian, D.D. and Zhou, J.H. (2019) 'Application of business intelligence based on big data in e-commerce data analysis', *E-Business Journal*, Vol. 15, No. 04, pp.29–30+96.

Sarker, I.H. (2019) 'Mining individualized context-dependent behavioral rules from smartphone data', *Journal of Ambient Intelligence and Smart Environments*, Vol. 11, No. 4, pp.369–370.

Shuvo, S.B., Ghosh, J. and Oyshi, A.S. (2019) 'A data mining based approach to predict autism spectrum disorder considering behavioral attributes', *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.

Sotomayor-Bonilla, J., Callejo-Canal, E.D., González-Salazar, C. et al. (2021) 'Using data mining and network analysis to infer arboviral dynamics: the case of mosquito-borne flaviviruses reported in Mexico', *Insects*, Vol. 12, No. 5, pp.398–409.

Sun, D., Leurent, F. and Xie, X. (2020) 'Floating car data mining: identifying vehicle types on the basis of daily usage patterns', *Transportation Research Procedia*, Vol. 47, No. 21, pp.147–154.

Wei, J. and Wang, Y. (2019) 'Analysis and research on e-commerce user behavior based on big data of online shopping platform', *Journal of Hubei Polytechnic University*, Vol. 35, No. 03, pp.34–38.

Wu, X.X., Huang, X.L., Lin, X.R., Li, Y.L. and Huang, Z.M. (2020) 'Data mining the prognostic biomarkers: miRNAs targeting the chemotherapeutic agent 5-fluorouracil in colon cancer', *Clinical Laboratory*, Vol. 66, No. 3, pp.69–75.

Zhao, Z.H. and Xu, Z.H. (2019) 'Short-term forecast of commodity sales volume based on e-commerce network data mining', *Logistics Sci-Tech.*, Vol. 42, No. 08, pp.1–7, p.12.

Ziaei-Bideh, A. and Namakshenas-Jahromi, M. (2020) 'Profiling green consumers with data mining', *Chapters*, Vol. 41, No. 32, pp.55–64.