Detection method of e-commerce cluster consumption behaviour based on data feature mining

Ming Yang

School of Business,
Chongqing College of Electronic Engineering,
Chongqing 401331, China
Email: 270562627@qq.com

Abstract: In order to effectively improve the accuracy and efficiency of e-commerce cluster consumption behaviour detection, an e-commerce cluster consumption behaviour detection method based on data feature mining is proposed. The concept and process of data feature mining and the e-commerce cluster consumption behaviour are analysed, and the characteristics of the e-commerce cluster consumption behaviour data with multiple characteristics are extracted. The Laplace feature mapping method is used to pre-process the extracted data features of e-commerce cluster consumption behaviour, the cyclic neural network structure is used to classify the data of e-commerce cluster consumption behaviour, and the data feature mining method is used to construct the detection model of e-commerce cluster consumption behaviour, so as to realise the detection of e-commerce cluster consumption behaviour. Experimental results show that the proposed method can effectively improve the detection accuracy and efficiency of e-commerce cluster consumption behaviour.

Keywords: data feature mining; cyclic neural network; Laplace feature mapping; e-commerce clustering; consumer behaviour detection.

Reference to this paper should be made as follows: Yang, M. (2023) 'Detection method of e-commerce cluster consumption behaviour based on data feature mining', *Int. J. Reasoning-based Intelligent Systems*, Vol. 15, No. 1, pp.29–34.

Biographical notes: Ming Yang graduated from the Guizhou University with a Master's in Business Administration in 2014, and graduated with a Master's in Engineering from the Chongqing University of Posts and Telecommunications in 2015. Now, he is studying for a PhD degree at the University of Uttar Pradesh, Malaysia. He is currently an Associate Professor at the School of Business of Chongqing College of Electronic Engineering. His research interests include e-commerce, computer technology and other fields.

1 Introduction

At present, with the rapid development of e-commerce, online shopping has gradually become one of the main consumption modes. Compared with offline shopping, online shopping has the advantages of low price, wide variety and less affected by business hours and regions (Chen, 2020; Tezza and Borba, 2021; Zhang et al., 2018). However, it is precisely because of the huge commodity information and the diversity of shopping that consumers often spend a lot of time and energy when looking for products suitable for themselves. At the same time, in the e-commerce platform, due to the increasing market competition, businesses constantly refine the needs of commodities in meeting the actual needs of customers, so as to narrow the audience of each commodity (Vera-Baquero et al., 2021; Kim et al., 2021). In the future competition and development process, how to quickly and efficiently find suitable consumer groups and formulate more targeted market strategies is a very key problem. In addition, in the e-commerce platform, consumers will produce a large number of cluster consumption behaviours, which can analyse consumers' shopping intentions and consumption habits, and make one-to-one accurate recommendations.

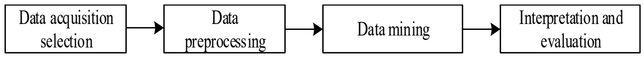
At present, scholars in related fields have studied the detection of consumption behaviour. Kao et al. (2021) proposed an offline consumer behaviour detection method. This study attempts to apply the characteristics of online channel to physical channel by using image object tracking and image detection technology. Through this inclusion, physical channels can provide consumers with more favourable experience and interaction, and physical store owners can more accurately understand the consumption behaviour of store consumers. The information obtained through the system can be provided to the shopkeeper as a reference for commodity placement, shelf layout and consumer circulation path planning. This study uses image processing technology to locate the region of interest, and uses target tracking to obtain the consumer trajectory, which successfully realises the consumer tracking characteristics of the physical channel online platform, while retaining the unique experience of the physical channel. However, this method has the problem of insufficient detection accuracy. Liashenko et al. (2021) proposed a consumer behaviour detection method in food retail chain based on machine learning algorithm. In the course of this study, a survey was created to determine the behaviour patterns of modern consumers according to their selection criteria and their responses to questions based on behavioural economics theorems. The results are monitored by machine learning algorithm, and then the random forest classification algorithm is trained. According to the results of contour analysis, K-means clustering is selected as the main detection for further modelling. This study provides a tool for customer preference classification and analysis of current industry trends, but the overall detection of this method takes a long time. Lin and Wang (2022) propose a consumption behaviour detection method based on self encoder isolated forest. This method reduces the dimension of e-consumer shopping data through the automatic encoder, detects the behaviour by using isolated forest, and optimises the parameters by using grid search algorithm. According to the results of parameter optimisation, a detection model is constructed to complete the detection of e-consumption behaviour. However, the detection accuracy of this method still needs to be further improved.

In order to solve the above problems, this paper studies the detection method of e-commerce cluster consumption behaviour by using data mining technology. Based on the data feature mining method, the data features of e-commerce cluster consumption behaviour are extracted. The Laplace feature mapping method is used to pre-process the extracted features of e-commerce cluster consumption behaviour data. The cyclic neural network structure is used to classify the e-commerce cluster consumption behaviour data, and the e-commerce cluster consumption behaviour detection model is constructed to realise the e-commerce cluster consumption behaviour detection. This method has good detection effect, high detection accuracy and efficiency of e-commerce cluster consumption behaviour.

2 Concept and process of data feature mining

Data mining is to extract and mine valuable knowledge of models and rules from massive data (Mao, 2018; Wang et al., 2021). A large amount of data in the database often contains more advanced information, such as rules, judgements, etc. these advanced information can only be obtained by structured processing. Data mining technology finds potential patterns that may affect information detection and decision-making through analysis, deduction and reasoning, so as to build a new data model, which is helpful for decision-makers to make correct decisions. Data mining is an important part of knowledge discovery. It includes four parts: data collection, pre-processing, mining and evaluation, as shown in Figure 1.

Figure 1 Data feature mining process



Data mining is an important part of knowledge discovery. It is the repetition of one or more stages. The original data actually collected are chaotic, repetitive and incomplete. Data collection and selection is to identify a set of data to be analysed, so as to reduce the scope of data processing. Data pre-processing mainly includes two parts: data cleaning and feature subset selection. Its function is to process unclean data and filter the feature subset (Guo and Zhang, 2019; Fan et al., 2018). In the process of data mining, the actual mining operations are analysed, and the corresponding mining methods are selected. A reasonable expression of the excavation results is a reasonable explanation.

3 Detection method of e-commerce cluster consumption behaviour

3.1 *Extract data characteristics of e-commerce cluster consumption behaviour*

Because there are many types of data features describing the e-commerce cluster consumption behaviour, the e-commerce cluster consumption behaviour data with multiple types of features should be extracted. Here, based on the e-commerce cluster consumption behaviour, the feature extraction of e-commerce cluster consumption behaviour data is realised to improve the accuracy of e-commerce cluster consumption behaviour detection (Lei et al., 2020). The specific implementation process is as follows:

The consumption data types of e-commerce clusters mainly include: browsing goods, consumption times, consumption amount, consumption preference, consumption cycle data, product purchase volume, etc. In this study, 1,000 cluster consumption data were randomly collected to meet the requirements of data feature extraction. The specific reflection of e-commerce cluster consumption behaviour can be realised through the e-commerce cluster consumption behaviour data set $H_1, H_2, H_3, \ldots, H_M$. It is a common data feature extraction method to directly extract the local e-commerce cluster consumption behaviour data features, and obtain the feature subset H_1 after processing the e-commerce cluster consumption behaviour data. The calculation formula is:

$$H_1 = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & \cdots & h_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ h_{B1} & h_{B2} & \cdots & h_{BM} \end{bmatrix} \quad (1)$$

By analysing the e-commerce cluster consumption behaviour, the characteristic quantity of e-commerce cluster consumption behaviour is selected as the feature subset, and the data characteristics of e-commerce cluster consumption behaviour are extracted.

3.2 Pre-processing the data characteristics of e-commerce cluster consumption behaviour

In this paper, Laplace feature mapping method is used to pre-process the extracted data features of e-commerce cluster consumption behaviour (Chen et al., 2019; Ke Mppainen et al., 2021). This method can construct the relationship between the data characteristics of e-commerce cluster consumption behaviour from a local perspective. Laplace feature mapping method reconstructs the local structural features of e-commerce cluster consumption behaviour data by establishing a similarity graph. The feature pre-processing method of e-commerce cluster consumption behaviour data based on Laplacian feature mapping believes that if instances i and j are very similar in high-dimensional space, then in the pre-processed target low-dimensional subspace, instance i and j should be as close as possible and retain similarity.

Let n represent the total number of data features of e-commerce cluster consumption behaviour, m represents the spatial dimension, Y is the matrix of $n \times m$, and the row vector y_i^T is used to represent the vector i of data features in the m dimensional subspace of the object, while the objective function required for Laplace feature mapping is:

$$X = \min \sum_{i,j} W_{i,j} \times \left\| Y - y_i^T \right\|^{m \times n} \tag{2}$$

In formula (2), $W_{i,j}$ describes the graph Laplacian matrix, and Y represents a generalised eigenvalue with m smallest eigenvalues.

A domain graph is established by Laplace. The nodes on the graph correspond to each data feature point, and the relationship between nodes is judged by the proximity of adjacent points. The effect of the process can be described by the following formula:

$$M^D \Rightarrow M^L, (L \le D) \tag{3}$$

In formula (3), M^L represents the mapping characteristic in the L dimensional space, and M^D represents the initial characteristic of the D dimensional space. A weighted graph is built according to the p points $z_1, z_2, \ldots, z_k$ existing in M^D. Among them, the adjacent points of each group are connected with the edge of each point, and a connecting edge is set between each group of adjacent nodes i and j. The genetic algorithm is used to calculate the characteristic weight of the e-commerce cluster consumption behaviour data, and the corresponding weight of the connecting edge is calculated. When the node i and j are connected, the following formula exists:

$$W_{i,j} = e - \frac{\left\| z_i - z_j \right\|^2}{4t} \tag{4}$$

On this basis, the mapping value corresponding to the generalised eigenvalue is calculated by using the domain graph:

$$Qy = \alpha Ay \tag{5}$$

In formula (5), Q represents the Laplace matrix, which can be calculated by the following formula:

$$Q = A - W \tag{6}$$

In formula (6), the element existing in A is the sum of each column in W, which describes the diagonal diagonal weight matrix. The calculation formula is as follows:

$$A_i = \sum_j W_i \tag{7}$$

Through the above process, the data feature pre-processing of e-commerce cluster consumption behaviour is completed.

3.3 Build a detection model of e-commerce cluster consumption behaviour

On the basis of pre-processing the data characteristics of e-commerce cluster consumption behaviour, the data feature mining method is used to construct the e-commerce cluster consumption behaviour detection model to realise the e-commerce cluster consumption behaviour detection.

First, the Nvs1 structure RNN is used to classify the consumption behaviour data of e-commerce clusters. At the end of data training, there will be a separate output value for detecting the consumption behaviour of e-commerce online shopping groups. The RNN classification is marked as:

$$E = \begin{cases} E_0 = 0 \\ E_1 = 1 \end{cases} \tag{8}$$

The data are graded according to the number of consumption behaviours of different online shopping groups. To a certain extent, the behaviour order of this level is as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1l} \\ r_{21} & r_{22} & \cdots & r_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{l2} & \cdots & r_{kl} \end{bmatrix} \tag{9}$$

In formula (9), l represents the length of e-commerce cluster consumption behaviour data, and k represents the total number of input e-commerce cluster consumption behaviour records. The hidden layer $u^0 = 0$ is initialised at t_1 time.

Input the input layer, hidden layer, and output layer parameters S, N, and C to construct an e-commerce cluster consumption behaviour detection model:

$$u^{(t)} = o\left(Sx_t + Nu_{t-1} + a\right) \tag{10}$$

In formula (10), o is the activation function of the RNN, and a is the bias term of the linear relationship of the e-commerce cluster consumption behaviour detection model. In order to obtain a single output result, only the consumption behaviour of the last e-commerce online shopping group is output and converted. Because this is a classification problem, the *softmax* function is used, and finally:

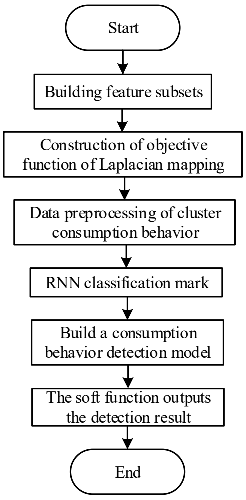
$$score = softmax\left(Cu_t + s\right) \tag{11}$$

In formula (11), s is the value obtained from the previous hidden state. d^{t+1} is the final loss generated by the RNN, which is used to measure the distance between the output f^{t+1} and the training result g^{t+1}, and to judge the detection effect of the e-commerce cluster consumption behaviour detection model. The loss function of RNN selects the cross entropy function, and the definition formula is as follows:

$$d^t = -\hat{\beta}_t \log(\beta_t) \tag{12}$$

In formula (12), β_t represents the correct answer of the e-commerce cluster consumption behaviour data, and $\hat{\beta}_t$ represents the detection output value of the e-commerce cluster consumption behaviour detection model. In this way, through the above process, the construction of the e-commerce cluster consumption behaviour detection model is completed, so as to realise the e-commerce cluster consumption behaviour detection. Its e-commerce cluster consumption behaviour detection process is shown in Figure 2.

Figure 2 Flow chart of e-commerce cluster consumption behaviour detection



4 Experimental simulation and analysis

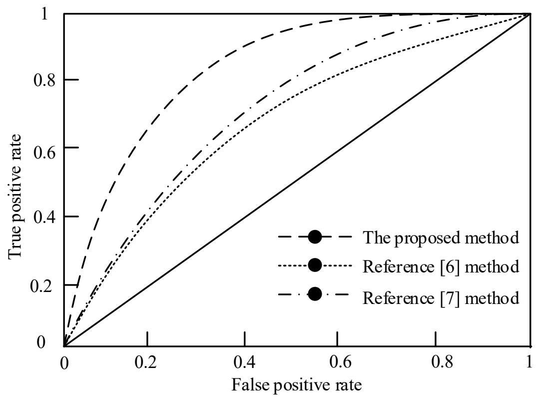
4.1 Experimental data and scheme

In order to verify the effectiveness of the detection method of e-commerce cluster consumption behaviour based on data feature mining, the data used in this paper is processed and filtered using Mysql5.6, and the data is stored as a csv file. The experiment uses Python3.5 programming, and uses the open source Anaconda4.2.0 to manage Python packages. The recurrent neural network in this paper adopts the TensorFlow framework, and randomly selects 1,000 e-commerce cluster consumption behaviour data as the experimental sample data. Taking the consumption behaviour detection effect, consumption behaviour detection accuracy and detection time as experimental comparison indexes, the e-commerce cluster consumption behaviour detection is carried out by using the Kao et al. (2021) method, Liashenko et al. (2021) method and the proposed method respectively to verify the effectiveness of the proposed method.

4.2 Comparison results of e-commerce cluster consumption behaviour detection results

In order to verify the detection effect of e-commerce cluster consumption behaviour of the proposed method, ROC curve is used as an evaluation index. The AUC value is determined by the area surrounded by ROC curve and coordinate axis. The larger the area, the better the detection effect of e-commerce cluster consumption behaviour of the representation method. Using the method of Kao et al. (2021), the method of Liashenko et al. (2021) and the proposed method to compare, the results of the comparison of the detection effect of different methods of e-commerce agglomeration consumption behaviour are shown in Figure 3.

Figure 3 Comparison results of e-commerce cluster consumption behaviour detection results of different methods



According to Figure 3, the maximum AUC value of the ROC curve area of the method of reference (Kao et al., 2021) is 0.765, the maximum AUC value of the ROC curve area of the method of Liashenko et al. (2021) is 0.782, and the maximum AUC value of the ROC curve area of the proposed method is as high as 0.916. It can be seen that, compared with the method of Kao et al. (2021) and the method of Liashenko et al. (2021), the maximum AUC value of the ROC curve area of the proposed method is larger, indicating that the proposed method has a better effect on the detection of e-commerce agglomeration consumption behaviour.

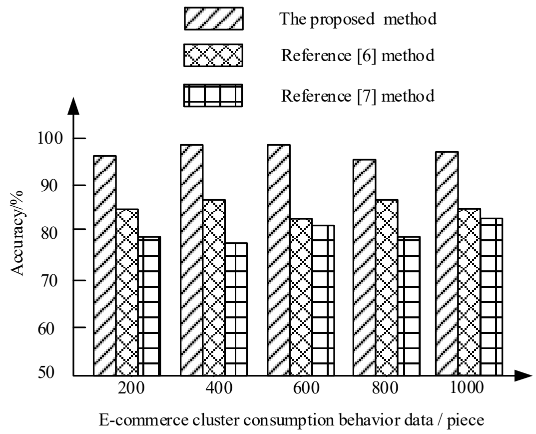
4.3 Comparison of detection accuracy of e-commerce cluster consumption behaviour

Further verify the detection accuracy of e-commerce cluster consumption behaviour of the proposed method, and take the accuracy as the evaluation index. The higher the accuracy, the higher the detection accuracy of e-commerce cluster consumption behaviour. The calculation formula is as follows:

$$J_Q = \frac{J_G}{J_Z} \times 100\% \tag{13}$$

In formula (13), J_G refers to the number of correctly detected e-commerce cluster consumption behaviour data, and J_Z refers to the total number of e-commerce cluster consumption behaviour data that needs to be detected. The method of Kao et al. (2021), the method of Liashenko et al. (2021) and the proposed method are used to compare, and the comparison results of the detection accuracy of e-commerce agglomeration consumption behaviour of different methods are obtained as shown in Figure 4.

Figure 4 Comparison results of detection accuracy of e-commerce cluster consumption behaviour by different methods



According to Figure 4, when there are 1,000 e-commerce cluster consumption behaviour data, the average e-commerce cluster consumption behaviour detection accuracy of the method of reference (Kao et al., 2021) is 86.2%, and the average e-commerce cluster consumption behaviour detection accuracy of the method of reference (Liashenko et al., 2021) is 79.8%. The average detection accuracy of e-commerce cluster consumption behaviour of the proposed method is as high as 97.1%. Therefore, the proposed method has high detection accuracy of e-commerce cluster consumption behaviour, and can effectively improve the detection accuracy of e-commerce cluster consumption behaviour.

4.4 *Comparison of detection efficiency of e-commerce cluster consumption behaviour*

On this basis, further verify the detection efficiency of e-commerce cluster consumption behaviour, and take the detection time as the evaluation index. The shorter the detection time, the higher the detection efficiency of e-commerce cluster consumption behaviour. By comparing the method of Kao et al. (2021), the method of Liashenko et al. (2021) and the proposed methods, we get the comparison results of the detection time of e-commerce cluster consumption behaviour of different methods, as shown in Table 1.

Table 1 Comparison results of detection time of e-commerce cluster consumption behaviour by different methods

E-commerce cluster consumption behaviour data /piece	*The proposed method/s*	*The method of Kao et al. (2021)/s*	*The method of Liashenko et al. (2021)/s*
200	7.2	8.1	10.8
400	9.1	12.9	14.3
600	11.9	16.2	18.5
800	14.3	19.8	22.4
1,000	17.2	23.6	27.9

According to Table 1, with the increase of e-commerce cluster consumption behaviour data, the detection time of e-commerce cluster consumption behaviour of different methods increases. When there are 1,000 e-commerce cluster consumption behaviour data, the e-commerce cluster consumption behaviour detection time of the method of reference (Kao et al. 2021) is 23.6 s, the e-commerce cluster consumption behaviour detection time of the method of reference (Liashenko et al., 2021) is 27.9 s, while the e-commerce cluster consumption behaviour detection time of the proposed method is only 27.9 s. Therefore, the detection time of e-commerce cluster consumption behaviour of the proposed method is short, which can effectively improve the detection efficiency of e-commerce cluster consumption behaviour.

5 Conclusions

The detection method of e-commerce cluster consumption behaviour based on data feature mining is proposed in this paper. The data of e-commerce cluster consumption behaviour is detected by using data feature mining method and cyclic neural network. This method has a good effect on the detection of e-commerce cluster consumption behaviour, and can effectively improve the detection accuracy and efficiency of e-commerce cluster consumption behaviour, with an average detection accuracy of 97.1%. However, this method does not take into account the interspersed behaviour of e-commerce cluster consumers among different categories of goods, and ignores the relationship between e-commerce product categories. Therefore, in the following research, we can improve the data dimension of e-commerce cluster consumption behaviour, so as to further optimise the data detection effect.

References

Chen, X., Fang, S., Li, Y. and Wang, H. (2019) 'Does identification influence continuous e-commerce consumption? The mediating role of intrinsic motivations', *Sustainability*, Vol. 11, No. 7, pp.1–20.

Chen, Y. (2020) 'Research on personalized recommendation algorithm based on user preference in mobile e-commerce', *Information Systems and e-Business Management*, Vol. 18, No. 4, pp.837–850.

Fan, J., Tang, L., Zhu, W. and Zou, B. (2018) 'The Alibaba effect: spatial consumption inequality and the welfare gains from e-commerce', *Journal of International Economics*, Vol. 114, No. 9, pp.203–220.

Guo, L. and Zhang, D. (2019) 'EC-structure: establishing consumption structure through mining e-commerce data to discover consumption upgrade', *Complexity*, Vol. 20, No. 9, pp.1–8.

Kao, Y.T., Lin, H.W. and Qiu, D.Y. (2021) 'Implementation of offline consumer behavior tracking', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 35, No. 9, pp.215–219.

Ke Mppainen, T., Frank, L., Makkonen, M. and Hyvnen, O.I. (2021) 'Barriers to responsible consumption in e-commerce: evidence from fashion shoppers', *Digital Support from Crisis to Progressive Change*.

Kim, G., Wang, W. and Ha, H.K. (2021) 'Pricing strategy for own shipping service of e-commerce platform using two-sided market theory', *Electronic Commerce Research and Applications*, Vol. 49, No. 2, pp.101–108.

Lei, F., Lei, S, Liu, J. and Wu, L. (2020) 'Studying customer behavior of shopping on e-commerce platforms from the perspective of college students', *2020 International Conference on E-Commerce and Internet Technology (ECIT)*, IEEE.

Liashenko, O., Kravets, T. and Prokopenko, M. (2021) 'Consumer behavior clustering of food retail chains by machine learning algorithms', *Access Journal*, Vol. 23, No. 3, pp.234–251.

Lin, G. and Wang, Y. (2022) 'Abnormal behavior detection of online shoppers based on autoencoder-isolation forest', *Computer Applications and Software*, Vol. 39, No. 2, pp.253–258.

Mao, Y. (2018) 'Electronic word-of-mouth for e-commerce consumption in mobile social network: a case study from WeChat', *WHICEB 2016 Proceedings*, Vol. 8, No. 2, pp.151634–151649.

Tezza, R. and Borba, E. (2021) 'Analysis of users' perceptions on mobile electronic commerce of fashion', *Revista Alcance*, Vol. 28, No. 2, pp.197–211.

Vera-Baquero, A., Phelan, O., Slowinski, P. and Hannon, J. (2021) 'Open source software as the main driver for evolving software systems toward a distributed and performant e-commerce platform: a Zalando fashion store case study', *IT Professional*, Vol. 23, No. 1, pp.34–41.

Wang, E.P., Ning, A.N., Geng, X.H., Gao, Z. and Kiprop, E. (2021) 'Consumers' willingness to pay for ethical consumption initiatives on e-commerce platforms', *Journal of Integrative Agriculture*, Vol. 20, No. 4, pp.1012–1020.

Zhang, M., Yao, Y., Jiang, Y., Li, B. and Tang, C. (2018) 'Accountable mobile e-commerce scheme in intelligent cloud system transactions', *Journal of Ambient Intelligence and Humanized Computing*, Vol. 9, No. 6, pp.1889–1899.