

Research Article

Brand Preference Prediction Method of Cross-Border E-Commerce Consumers Based on Potential Tag Mining

Xujie Qin 

School of Management, Anhui Business and Technology College, Hefei 231131, Anhui, China

Correspondence should be addressed to Xujie Qin; qinxj@ahbvc.edu.cn

Received 13 April 2022; Revised 21 May 2022; Accepted 30 May 2022; Published 24 June 2022

Academic Editor: Wei Liu

Copyright © 2022 Xujie Qin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are many brands in cross-border e-commerce platforms. Obtaining consumers' preference for brands will help promote the development of cross-border e-commerce industry. A brand preference prediction method of cross-border e-commerce consumers based on potential tag mining is proposed. Preprocess the cross-border e-commerce brand comment information obtained, build a HowNet emotion dictionary, and calculate consumers' emotional tendency towards the brand on this basis. The projection pursuit regression model is optimized by differential evolution algorithm to reduce the dimension of the obtained consumer brand emotion information. Mining the potential labels of the information after dimensionality reduction, combined with Bayesian personalized sorting method and paired interaction tensor decomposition method, this paper constructs a brand preference's prediction model to predict the brand preference of cross-border e-commerce consumers. The experimental results show that the proposed method has high accuracy of brand tendency calculation results, small average absolute error of prediction results, and high model accuracy.

1. Introduction

Under the background of the rapid development of Internet technology, cross-border e-commerce has been more restricted in the international market. With the fierce trend of mutual suppression, e-commerce not only has the problem of product infringement, but also reduces profits [1, 2]. In the international mainstream market, international big brands occupy an important position, so domestic e-commerce businesses mainly focus on nonmainstream products. The low degree of consumers' satisfaction and preference for domestic brands is mainly due to a series of problems in China's cross-border e-commerce, including low internationalization level, low market position, and poor innovation ability [3]. In order to improve the position of China's cross-border e-commerce in the international market, it is necessary to obtain consumers' preference for brands.

Li and Maomao [4] and others designed a consumer preference questionnaire to obtain the current situation information of consumer brand preference, extracted brand

features from four aspects of society, brand, service, and price perceived value through in-depth learning model, and established a consumer brand preference prediction model according to the feature extraction results. However, this method cannot accurately calculate consumer brand preference, which has the problem of large average absolute error of prediction results. Sun et al. [5] and others obtained online text and consumer keyword index data and used sliding time window technology to complete the above data screening. At the same time, combined with a variety of machine learning methods, the prediction model of consumer information index is established, to obtain the consumer information index of different brands and realize the prediction of consumer brand preference. The disadvantage is that the accuracy is low.

In order to solve the problems in the above methods based on the calculation of brand preference emotional tendency, this paper realizes the prediction of cross-border e-commerce consumers' brand preference by mining potential tags.

2. Calculation of Brand Preference and Emotional Tendency

2.1. Comment Pretreatment. One of the main tasks to predict the brand preference of cross-border e-commerce consumers is comment preprocessing. The specific process is as follows in Figure 1.

Reprocess the comment data according to the process in Figure 1 to ensure the accuracy of brand preference and emotional tendency prediction. Obtain brand comment data from the Internet or other websites through web crawler technology. The comment data is segmented by means of annotation text, word segmentation, and annotation, which is divided into clauses with the smallest unit to ensure the integrity of its semantic expression.

Compared with English writing methods, Chinese writing methods are usually quite different. In Chinese writing rules, words in sentences are not divided by display separators such as spaces. The smallest semantic unit in the text is words, which is of great significance for emotion analysis about consumer brand. As the most basic step, word segmentation has many research results in the practical problems of Chinese natural language processing [6, 7]. The function of words in sentences can be described by part of speech. Therefore, part of speech tagging is an important link in the process of comment preprocessing. The commonly used open source tools for part of speech tagging and Chinese word segmentation are shown in Table 1.

In natural language processing, the word segmentation package in the tool library is usually used to complete data segmentation, and then the weight of different data is given according to the unused work and emotional needs to build the final comment database and emotion table. So far, comment data preprocessing is realized.

2.2. Construction of Emotional Dictionary. In terms of natural language processing, most countries have built knowledge bases, such as HowNet in China, Korean WordNet in Korea, and mindnet in Microsoft. HowNet has a high position in Chinese emotional dictionary. HowNet contains the relationships between ideas. It is a semantic web based on world knowledge. Its content is more detailed. The main relationships in HowNet are shown in Figure 2.

The concept in HowNet shown in Figure 2 includes many semantic relationships such as synonymy and antonymy. Among them, the most important relationship is the synonymous and antisense relationship of words. According to the relationship between different words, the word attributes of corresponding words can be decomposed.

Decompose concepts in HowNet to obtain multiple semaphores. The similarity $S(w_1, w_2)$ between words can be calculated through the distance $d(w_1, w_2)$ value in the semaphore tree, as shown in formula (1):

$$S(w_1, w_2) = \frac{\beta}{\beta + d(w_1, w_2)}, \quad (1)$$

where β represents the change parameter, which represents the polarity value of this type of emotion, and this parameter

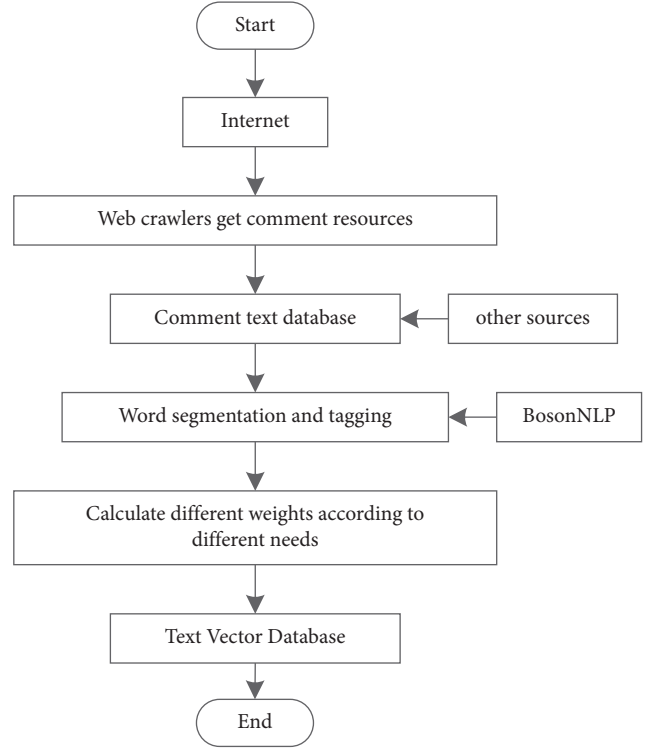


FIGURE 1: Comment preprocessing.

is greater than zero. The calculation results of the above formula are prone to errors. Therefore, the proposed method calculates the similarity between words according to the correlation between the word and the negative word reference word b_1 and the positive word reference word b_2 , that is, the semantic tendency value $t(w)$, as shown in formula (2):

$$t(w) = \frac{\sum_{i=1}^n s(w_i, b_1)}{n} - \frac{\sum_{j=1}^m s(w_j, b_2)}{m}, \quad (2)$$

where n and m represent the number of positive and negative reference words, respectively. When the semantic propensity value $t(w)$ of a word is positive, it belongs to a negative word. When the semantic propensity value $t(w)$ of a word is negative, it belongs to a positive word. Specific subjects based on brand preference classification fill the existing affective polarity dictionary, and the cross-border e-commerce prediction method of the consumer brand preference based on potential tag mining describes the attributes of words through binary W , as shown in formula (3):

$$W = (D, V), \quad (3)$$

where D represents the emotional polarity corresponding to the word, and its value is 1 or -1 . When the value is -1 , it belongs to negative words, and when the value is 1, it belongs to positive words. The nouns and adjectives of B and V are as shown in formula (4):

$$V = (B, S, M). \quad (4)$$

According to the attributes of two tuple descriptors, the emotion dictionary is constructed, which is used to judge the emotional bias in brand comments.

TABLE 1: The commonly used open source tools for part of speech tagging and Chinese word segmentation.

Analysis tools	Part of speech tagging	Analysis granularity	Accuracy	Interface	Website
Pao ding-analysis	None	Multiple choice	—	Jar package	https://code.google.com/p/paoding/
BosonNLP	Having	Multiple choice	0.947	REST API	https://bosonnlp.com/dev/center
IK analyzer	None	Large	0.748	Jar package	https://www.oschina.net/p/ikanalyzer
Pan gu segment	None	Multiple choice	0.803	None	https://pangusegment.codeplex.com/
NLPIR	Having	Minor	0.913	Jar package	https://ictclas.nlpir.org/docs

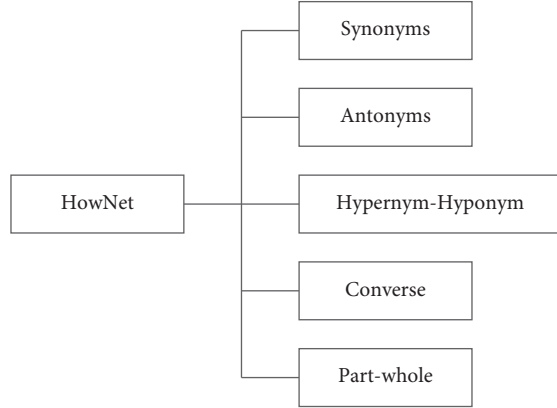


FIGURE 2: Main relationship diagram.

2.3. Calculation of Brand Emotional Tendency. The proposed method transforms the text content into vector operation in vector space through vector space model [8, 9], and the semantic similarity is measured by vector similarity.

In the process of text processing, vector space model uses multidimensional feature vectors to represent each document. When the number of documents in the data set is m , $m \times n$ dimensional vector space is used to represent the whole data set and abstract the text into vectors. The prediction method about brand preference of cross-border e-commerce consumers based on potential tag mining obtains text features through linear combination feature algorithm.

2.3.1. Word Frequency Method. The importance of features can be measured by word frequency [10, 11]. Count the frequency $G(t)$ corresponding to all words after word segmentation, delete the entries with $G(t)$ value of 1, and sort the remaining entries according to word frequency in order from large to small. The importance of words is directly proportional to word frequency.

2.3.2. Mutual Information Method. Mutual information describes the amount of information generated by mutual influence when two things happen at the same time [12, 13]. From the granularity of emotional analysis, the initial emotional analysis is to analyze the words with emotional color; that is, the words with emotional color are divided into positive and negative aspects, because the function of emotional dictionary is mainly to judge whether there are emotional words in the constructed emotional dictionary after word segmentation of a brand comment. \dot{D} represents the dictionary of emotional polarity, and its expression is as follows in formula (5):

$$\dot{D} = PF \cup NF, \quad (5)$$

where set NF is composed of negative words, set PF is composed of positive words, and set F is composed of words with frequency higher than one. The candidate word set J is obtained through the above process, as shown in formula (6):

$$J = \dot{D} \cap F. \quad (6)$$

There are n words in set J where the cross-border e-commerce consumer brand preference prediction method based on potential tag mining selects features through the mutual information method [14, 15] and obtains the mutual information $O(y, v)$ between category v and entry y through the following formula, as shown in formula (7):

$$O(y, v) = \log_2 \frac{P(y|v)}{P(y)}, \quad (7)$$

where $P(y)$ represents the probability of the number of documents to which entry y belongs in the total number of documents; $P(y|v)$ represents the probability that the entry y appears in the data of category C , and the calculation results are sorted in the order from large to small. For features, the average value $O(y)$ corresponding to the entry is calculated by the following formula:

$$o(y) = \sum_{i=1}^k O(v_i)O(y, v_i), \quad (8)$$

where k represents the number of categories; $O(v_i)$ represents the probability corresponding to category v_i .

2.3.3. Linear Combination Feature Selection. The proposed method combines word frequency and mutual information to calculate the feature score of consumer brand emotion, as shown in formula (9):

$$\text{Score} = (1 - \mu)O(y_i) + \mu \text{Freq}(y_i), \quad (9)$$

where μ represents the regulating factor; $\text{Freq}(y_i)$ represents the word frequency corresponding to the entry y_i . According to the brand emotion characteristics extracted in the above process, the brand emotion tendency is obtained.

3. Cross-Border E-Commerce Consumer Brand Preference Prediction Regression Algorithm

3.1. Projection Pursuit Regression. Projection pursuit combines computer science, modern statistics, and applied mathematics. It is a high-tech and can complete nonlinear,

nonnormal, and high-dimensional data processing and analysis [16, 17]. The proposed method constructs a projection pursuit regression model to process the brand emotional tendency data obtained in the above process, to realize the prediction of consumer brand preference.

3.2. Differential Evolution Algorithm. Different from evolutionary strategy, evolutionary programming, and genetic algorithm, differential evolutionary algorithm [18, 19] has many advantages. The algorithm obtains mutation operator according to the difference vector between parent individuals, sets probability, cross processes mutation individuals, and parent individuals, generates new individuals, that is, test individuals, and selects the optimal individual between test individuals and parent individuals according to the fitness. The prediction method about brand preference of cross-border e-commerce consumers based on potential tag mining adopts differential evolution algorithm to optimize the prediction method about brand preference to complete the dimensionality reduction processing of brand emotional tendency data. The processing process is as follows.

Considering that the brand affective tendency data obtained by projection pursuit regression has certain pertinence, scale, and representativeness, in order to obtain the optimal brand affective tendency data that meets the requirements, this paper uses differential evolution algorithm to reduce its dimension. Initialize the population in the feasible solution space, and a new population will be generated after the cross mutation operation of the current population. Then, the selection operation based on greedy idea is used to select the two populations one-to-one, so as to obtain a new generation of population. The specific process is to perform mutation operation on individual x_i^t in each t time through formula (9) to obtain the corresponding mutant individual v_i^{t+1} , as shown in formula (10):

$$v_i^{t+1} = x_{r_1}^t + K(x_{r_2}^t - x_{r_3}^t), \quad (10)$$

where $x_{r_1}^t$ is the parent basis vector, $(x_{r_2}^t - x_{r_3}^t)$ is the parent difference vector, and K is the scaling factor. Then, x_i^t and v_i^{t+1} were processed by cross operation to obtain test individual u_i^{t+1} . The objective functions of individuals u_i^{t+1} and x_i^t are constructed, and the individuals with lower function values are selected as the individuals of the new species group, as shown in formula (11).

$$x_i^{t+1} = \begin{cases} u_i^{t+1} & f(u_i^{t+1}) < f(x_i^t) \\ x_i^t & \text{else} \end{cases}, \quad (11)$$

where $f(\cdot)$ is the objective function. In order to reduce the complexity of the algorithm, the feasible set scale is N_1 , the infeasible group scale is N_2 , and the maximum scale is N_3 . The total complexity of the diversity group is $M = (N_1, N_3)$. Then the complexity of the difference algorithm after one iteration can be expressed as formula (12):

$$M = K + O(v_i) \lg(N_1 + N_3). \quad (12)$$

If the search scope is $N = N_1 + N_2 + N_3$, it indicates that the data dimension of brand emotional tendency is effectively reduced.

3.3. Regression Algorithm of Brand Preference Prediction Based on Potential Tag Mining. Based on potential tag mining, the proposed method combines Bayesian personalized ranking method BPR and paired interaction tensor decomposition method (PITF) to build a brand preference prediction model.

Tag addition belongs to the user's information release behavior, which often includes the user's understanding of the content characteristics and themes of the brand they are browsing or the expression of the content themes they are interested in, which more clearly implies that the user is very interested in the content themes involved in the brand. So, according to the contextual relationship between brands and users, the proposed method mines high-value label samples to obtain the following labels:

- (1) Negative class label: $T_B = \{t_B | (u, i, t_B) \notin D \cap (u, i, t_B) \notin D_p\}$.
- (2) Potential label: $T_P = \{t_P | (u, i, t_P) \in D_p\}$.
- (3) Positive label: $T_A = \{t_A | (u, i, t_A) \in D\}$.

On the basis of the above labels, make word frequency statistics on the brand preferred words of cross-border e-commerce consumers, arrange the word frequency in descending order, select the first 100 words, and draw the word cloud diagram using word cloud in word cloud module, as shown in Figure 3.

Based on the above labels, we can predict the brand preference of cross-border e-commerce consumers. Bayesian personalized ranking method (BPR) obtains the correct ranking of brands based on Bayesian theory [20], as shown in formula (13):

$$a(\Gamma | \chi_{u,i}) \propto a(\chi_{u,i} | \Gamma) a(\Gamma), \quad (13)$$

where $a(\Gamma)$ represents a priori probability function; Γ represent the parameters of the interactive tensor decomposition model; $a(x, u, i | \Gamma)$ represents the likelihood function. Assuming that all preferences (u, i) are independent of each other, the maximum a posteriori probability can be described by the following formula:

$$\max \prod_{(u,i) \in D} a(y_A \chi_{u,i} y_B | \Gamma) a(\Gamma), \quad (14)$$

where y_A represents the observation label; D represents the triple set of $\langle \text{user } u, \text{brand } i, \text{label } y_A \rangle$, and its expression is as follows in formula (15):

$$D = \{(u, i, y_A) | u \in U, i \in I, y_A \in Y\}, \quad (15)$$

$a(y_A \chi_{u,i} y_B | \Gamma)$ represents the probability of partial order relationship. The proposed method describes the above probability through sigmoid function, as shown in formula (16):

$$a(y_A \chi_{u,i} y_B | \Gamma) = \varsigma [G_A(\Gamma) - G_B(\Gamma)]. \quad (16)$$

In the formula, G_A and G_B represent the prediction score, and the parameters ς to be obtained can be calculated by the following formula:

$$\varsigma = \frac{1}{1 + \exp(-x)}. \quad (17)$$

According to the relationship among users, brands, and labels, the proposed method models the preference through the paired interaction tensor decomposition method to obtain the preference value of preference (u, i) for label y_A , as shown in formula (18):

$$G(\Gamma) = U_i V_i + U_u T_{y_A}^U + V_i T_{y_A}^V. \quad (18)$$

Among them, $U_i, V_i, U_u, T_{y_A}^U$, and $T_{y_A}^V$ all represent the implicit factor matrix.

Based on the above calculation results, build a prediction model of cross-border e-commerce consumer brand preference, as shown in formula (19):

$$\min_{\Gamma} Z_1 = \sum_D \ln \varsigma [G_A(\Gamma) - G_B(\Gamma)] + \frac{\varsigma}{2} \|\Gamma\|_2^2. \quad (19)$$

According to the solution results of the prediction model, the brand preference prediction of cross-border e-commerce consumers based on potential tag mining is completed.

4. Experimental Analysis

In order to verify the effectiveness of the brand preference prediction method of cross-border e-commerce consumers based on potential tag mining, the brand preference prediction method of cross-border e-commerce consumers based on potential tag mining, literature [4] method, and literature [5] method are used to predict their brand preference.

4.1. Data Analysis. This paper selects the shopping comment information of 200 consumers on Amazon's cross-border e-commerce platform as the data source and uses Python web crawler technology to capture the comment data of different brands. The 200 consumers, including college students and company employees, have a certain understanding of cross-border e-commerce brands. The specific situation of 200 users is shown in Tables 2 and 3.

4.2. Analysis of Dimensionality Reduction Effect of Brand Emotional Tendency Data. The dimensionality reduction effect of brand affective tendency data is related to the response speed of the final prediction model. Therefore, through the response speed index analysis, this paper uses differential evolution algorithm to reduce the dimensionality of brand affective tendency data. The test results are shown in Figure 4.

It can be seen from the analysis of Figure 4 that the response speed of this method is always lower than 60 ms,

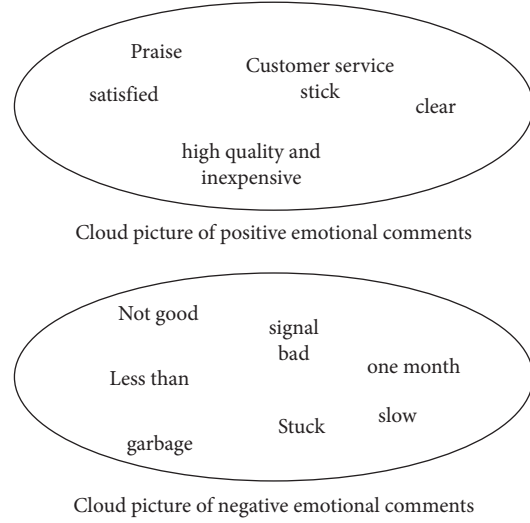


FIGURE 3: Cloud picture of words.

TABLE 2: Age distribution of consumers.

Age distribution	Number/person	Percentage/%
Under 20	58	29.0
21-30 years old	46	23.0
31-40 years old	69	34.5
41-50 years old	27	13.5
Over 50	0	0.0
Total	200	100.0

TABLE 3: Gender distribution of consumers.

Gender	Number/person	Percentage/%
Male	77	38.5
Female	123	61.5
Total	200	100.0

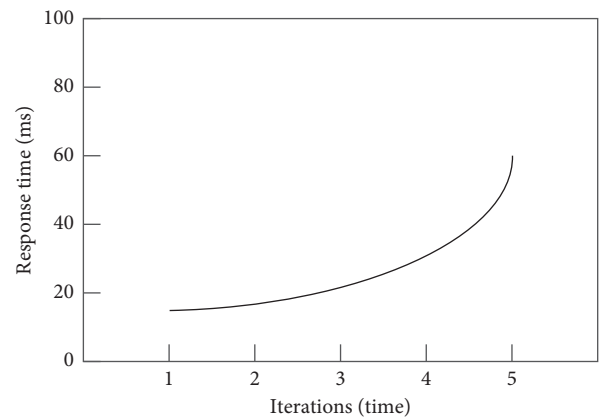


FIGURE 4: Response speed of this method.

which has a faster prediction effect of brand emotional tendency. The main reason is that this method uses differential evolution algorithm to reduce the dimension of the data on the basis of constructing the brand emotional tendency data analysis model based on projection pursuit

TABLE 4: Calculation results of consumers' emotional tendency.

Iterations/ time	Male consumers/%			Female consumers/%		
	Literature [4] methods	Literature [5] methods	The method of this paper	Literature [4] methods	Literature [5] methods	The method of this paper
1	60	50	86	58	60	94
2	50	76	89	60	59	92
3	70	80	87	70	69	93
4	64	40	86	68	59	92
5	42	68	88	59	56	90

regression, reduces the complexity and dimension of the data through multiple iterations, and improves the prediction response speed.

4.3. Calculation of Brand Emotional Tendency. The proposed method, literature [4] method, and literature [5] method are used to calculate the brand emotion tendency of the above consumers, and the calculation accuracy of different methods is compared, as shown in formula (20):

$$\text{Accuracy} = \frac{N_Z}{N_{\text{all}}}, \quad (20)$$

where N_Z represents the number of correctly calculated samples; N_{all} represents the total number of samples. According to formula (19), the calculation results of consumers' emotional tendency under the application of the proposed method, literature [4] method, and literature [5] method are obtained, as shown in Table 4.

In order to better display the calculation results of emotional tendency of male and female consumers, Table 4 is drawn into Figure 5. The following figure fully shows the changes.

Figure 5 is comparison chart of calculation of consumer sentiment propensity; it can be seen from Figure 5 that the accuracy of the three methods in calculating the brand sentiment propensity of male consumers is lower than that of female consumers. The main reason may be that male consumers publish less shopping comments, and the brand sentiment propensity data obtained by the corresponding methods are relatively less. At the same time, male consumers are more rational than female consumers and more rational for comments, Therefore, it is difficult to distinguish their emotional tendencies, resulting in low accuracy. For male consumers, the accuracy of the proposed method is 89% and that of female consumers is 94%, while the accuracy of literature [4] method and literature [5] method is 70% and 70%, 80% and 69%, respectively. From a comprehensive point of view, whether male or female consumers, the calculation results of brand emotion tendency of the proposed method are better than those of [4] and [5], because the proposed method constructs a HowNet emotion dictionary, which is used to calculate consumers' brand emotion tendency and improve the accuracy of the calculation results.

It can be seen from Figures 6 and 7 that the calculation results of brand emotion tendency of both male and female consumers by the proposed method are better than those in

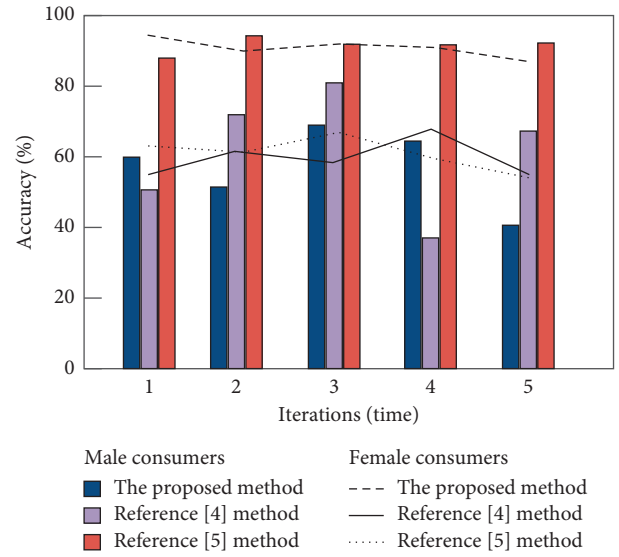


FIGURE 5: Comparison chart of calculation of consumers' emotional tendency.

[4] and [5], because the proposed method constructs a HowNet emotion dictionary to calculate consumers' brand emotion tendency on this basis, which improves the accuracy of the calculation results.

4.4. Prediction Results. The average absolute error MAE is used as an index to test the brand preference prediction results of the proposed method, the method in [4], and the method in [5], as shown in formula (21).

$$MAE = \frac{\sum_{i=1}^k |a_i - t_i|}{k}, \quad (21)$$

where a_i represents the prediction result; t_i represents the actual results; k represents the number of brands.

As can be seen from Figure 8, the average absolute error obtained by the proposed method when predicting consumers' brand preference for cross-border e-commerce consumers of different ages is low, with the highest error of 0.2%, which is lower than 0.6% of the method in [4] and 0.5% of the method in [5]. The main reason is that the proposed method obtains more accurate input data from the consumer comment data in the cross-border e-commerce platform through preprocessing and dimensionality reduction before prediction, the average absolute error of prediction is reduced, and the effectiveness of the proposed

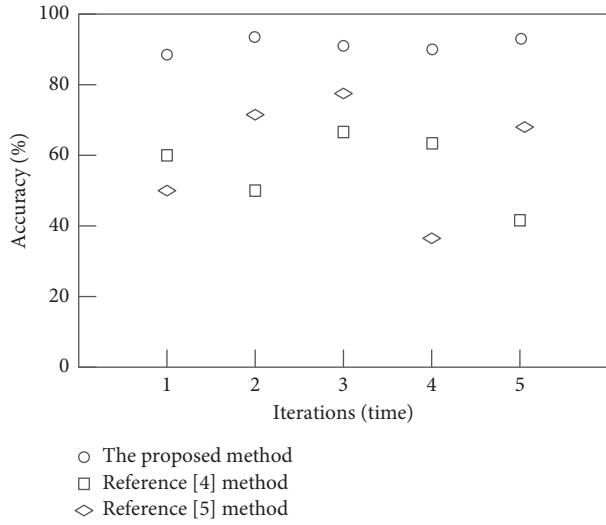


FIGURE 6: Calculation results of brand emotional tendency of male consumers.

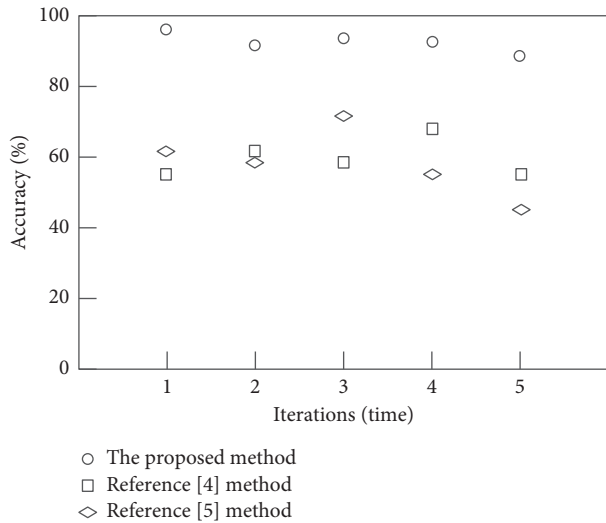


FIGURE 7: Calculation results of brand emotional tendency of female consumers.

method is verified. At the same time, according to Figure 7, the average absolute error of the prediction of the three methods is relatively high in the age group of 21 ~ 30. The main reason may be that the brands concerned by this group are more likely to buy goods for the parent group in addition to the acceptable brands in their own age group. At the same time, this group involves young couples with children, so their brands cover a wide range, resulting in a relatively high average absolute error of prediction.

4.5. Model Evaluation. Using the ROC curve, the farther the ROC curve from the pure opportunity line, the stronger the discrimination of the subjects, and test the accuracy of the model constructed by the proposed method, the method in [4], and the method in [5]. The test results are shown in Figure 9.

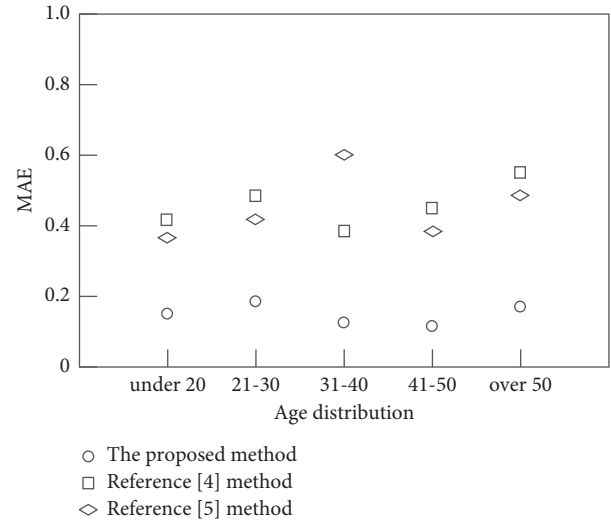


FIGURE 8: Prediction results of different methods.

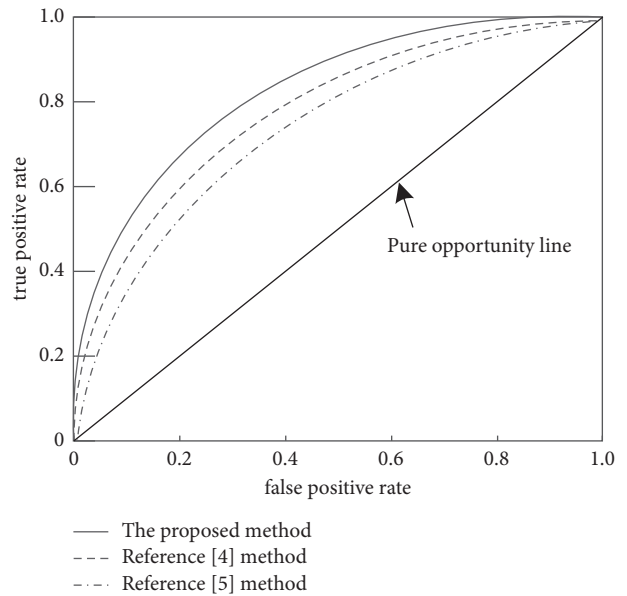


FIGURE 9: ROC curves of different methods.

The abscissa in Figure 9 is the false positive rate and the ordinate is the true positive rate. The model constructed by the proposed method is farther from the pure opportunity line, indicating that the accuracy of the model constructed by the proposed method is higher. The main reason is that the proposed method optimizes the model input and improves the prediction accuracy of the model by mining the potential labels of brand preference of cross-border e-commerce consumers.

5. Conclusions and Prospects

Aiming at the problems of low accuracy of brand tendency calculation results, large average absolute error of prediction results, and low model accuracy caused by many cross-border e-commerce brands, a cross-border e-commerce

consumer brand preference prediction method based on potential tag mining is proposed. Based on the calculation of cross-border e-commerce brand emotional tendency, this paper uses differential evolution algorithm to optimize the projection pursuit regression model, reduces the dimension of the obtained consumer brand emotional data, and realizes the prediction of brand preference based on Bayesian method. The experimental results show that this method has high accuracy for the calculation results of consumer brand tendency of different genders and different ages and improves the accuracy of the prediction model, It can provide data reference for the development of cross-border e-commerce industry, but the research on the prediction efficiency of the proposed method is insufficient, and improving the prediction efficiency will be taken as the research direction in future work.

Data Availability

The datasets generated for this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Key Research Project of Humanities and Social Sciences in Colleges and Universities of Anhui Province: Research on the Strategies of Transforming Cross-Border E-Commerce for Anhui Traditional Foreign Trade Enterprises from the Perspective of the "Belt and Road" Initiative (Project no. SK2018A0915), and Major Research Project of Humanities and Social Sciences in Colleges and Universities of Anhui Province: Research on the Policy and Path Of Low Carbon Economy Development in Anhui Province (Project no. SK2018ZD060).

References

- [1] X. Zhang and S. Liu, "Action mechanism and model of cross-border E-commerce green supply chain based on customer behavior," *Mathematical Problems in Engineering*, vol. 2021, no. 3, 11 pages, Article ID 6670308, 2021.
- [2] X. Zhang and Y. Shi, "Trade Facilitation and the Development of Cross-Border E-Commerce in China -- an Empirical Analysis Based on Double Difference Method," *East China Economic Management*, vol. 34, no. 2, pp. 94–103, 2020.
- [3] W. Su and Y. Wang, "Zhang Chonghui Economic effect of tax impact of foreign cross-border e-commerce: a study based on tso-dsge model," *Statistical research*, vol. 37, no. 1, pp. 17–32, 2020.
- [4] W. Li and C. Maomao, "Wang Weijun Research on online consumer preference prediction based on perceived value," *Journal of management*, vol. 18, no. 6, pp. 912–918, 2021.
- [5] J. Sun, J. Zhu, and H. Li, "Gao Zhe Research on rolling prediction of consumer confidence index based on Internet data," *Journal of Xi'an Jiaotong University (SOCIAL SCIENCE EDITION)*, vol. 41, no. 6, pp. 68–77, 2021.
- [6] Z. Cui, E. Zhao, and W. Luo, "Chinese word segmentation model of multi head attention for professional fields -- taking animal husbandry in Tibet as an example," *Chinese Journal of information technology*, vol. 35, no. 7, pp. 72–80, 2021.
- [7] F. Lizhou, G. Yang, and Xu Xue, "Two way matching Chinese word segmentation method based on n-gram," *Mathematical statistics and management*, vol. 39, no. 4, pp. 633–643, 2020.
- [8] Y. Wang and P. Wang, "Zhang Yu Research on technology information user portrait and scenario Service push based on vector space model," *Modern intelligence*, vol. 40, no. 2, pp. 3–10, 2020, + 25.
- [9] B. Xie, yuerun Dong, and L. pan, "Text emotion classification based on multi feature LSTM self attention," *Computer simulation*, vol. 38, no. 11, pp. 479–484, 2021.
- [10] X. Chen and D. Meurers, "Word frequency and readability: predicting the text-level readability with a lexical-level attribute," *Journal of Research in Reading*, vol. 41, no. 3, pp. 486–510, 2018.
- [11] W. Lu, S. Huang, J. Yang, Y. Bu, Q. Cheng, and Y. Huang, "Detecting research topic trends by author-defined keyword frequency," *Information Processing & Management*, vol. 58, no. 4, Article ID 102594, 2021.
- [12] M. Riahi-Madvar, A. Akbari Azirani, B. Nasersharif, and B. Raahemi, "A new density-based subspace selection method using mutual information for high dimensional outlier detection," *Knowledge-Based Systems*, vol. 216, no. 2, Article ID 106733, 2021.
- [13] J. Lewandowsky, S. J. Dongare, R. Martín Lima, M. Adrat, M. Schrammen, and P. Jax, "Genetic algorithms to maximize the relevant mutual information in communication receivers," *Electronics*, vol. 10, no. 12, p. 1434, 2021.
- [14] Q. Li, Y. Mou, and C. Zhanqi, "Defect Location Method Based on Mutual Information," *Design and Computer Engineering*, 2022.
- [15] F. Runze, X. Luo, and Z. Zhang, "Cigarette formula maintenance method based on mutual information of single cigarette," *Tobacco science and technology*, vol. 54, no. 3, pp. 65–71, 2021.
- [16] J. Yang and Q. Li, "Xie Jieyi Evaluation index system of regional scientific and technological Financial Development -- Analysis Based on projection pursuit model," *Research on science and technology management*, vol. 40, no. 6, pp. 69–74, 2020.
- [17] Y. Liu, "Shengxi Evaluation of urban high-quality development based on projection pursuit model weight optimization and its influencing factors," *Practice and understanding of mathematics*, vol. 51, no. 24, pp. 53–63, 2021.
- [18] C. Van Kien, H. P. H. Anh, and N. N. Son, "Adaptive inverse multilayer fuzzy control for uncertain nonlinear system optimizing with differential evolution algorithm," *Applied Intelligence*, vol. 51, no. 1, pp. 527–548, 2021.
- [19] Z. Pan, S. Fang, and H. Wang, "LightGBM technique and differential evolution algorithm-based multi-objective optimization design of DS-APMM," *IEEE Transactions on Energy Conversion*, vol. 36, no. 1, pp. 441–455, 2021.
- [20] C. Su, F. Ye, T. Liu, Y. Tian, and Z. Han, "Computation offloading in hierarchical multi-access edge computing based on contract theory and bayesian matching game," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13686–13701, 2020.