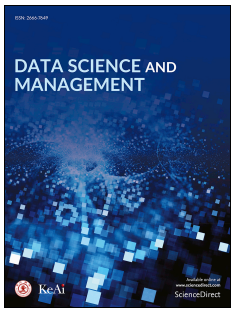
Journal Pre-proof



Application of user preference mining algorithms based on data mining and social behavior in brand building

Yuhan DONG

PII: S2666-7649(24)00019-5

DOI: https://doi.org/10.1016/j.dsm.2024.03.007

Reference: DSM 99

To appear in: *Data Science and Management*

Received Date: 2 September 2023

Revised Date: 25 March 2024

Accepted Date: 26 March 2024

Please cite this article as: DONG, Y., Application of user preference mining algorithms based on data mining and social behavior in brand building, *Data Science and Management*, https://doi.org/10.1016/j.dsm.2024.03.007.

Application of User Preference Mining Algorithms Based on Data Mining and Social Behaviour in Brand Building

Yuhan DONG[1*]

[1] Huhaisheng School of Public Administration, Nanyang Technological University, Singapore

*Corresponding Author: HUHA0008@e.ntu.edu.sg

Application of user preference mining algorithms based on data mining and social behavior in brand building

Abstract: Small and medium-sized enterprises currently suffer from a lack of branding. Therefore, to further promote their active branding, this study proposes a user preference mining algorithm based on data mining and social behavior. Employing this algorithm to study the degree of users' brand preference can provide data support for enterprises' brand building. The experimental results showed that the proposed algorithm outperforms previous algorithms in terms of performance, convergence, and accuracy. The area under the curve reached 0.953, indicating highly authentic output results with extremely high realism. In actual simulation experiments, its prediction results for the user's brand preference index are accurate, with an error of only 0.11, and the algorithm has extremely high ratings among industry insiders. In conclusion, the user-preference mining algorithm based on data mining and social behaviors suggested in this study plays a better role in promoting an enterprise's brand building. It can help the enterprise know the level of consumer preference for its brand; accordingly, it can determine the shortcomings in, provide effective and accurate data support for, and thereby promote its brand building.

Keywords: Mining algorithms; Brand building

1. Introduction

Building and applying brand value (BV) have become particularly critical tasks for small and micro firms (He and Calder, 2020). In recent years, the importance of an enterprise's brand has become increasingly apparent for its healthy development. Currently, although some businesses have excellent industrialized production capabilities, their BV still falls short of those of other well-known businesses (Kolbl et al, 2020). Businesses must develop BV through internal resources, and the fundamental issue facing today's businesses is the shortage of resources to increase economic efficiency through BV development (Guan et al, 2021). Users—or the performance of consumer preferences—can frequently indicate a consumer's recognition of or resistance to a certain brand (Findley et al, 2020), and the brand development (BD) of an organization can frequently be researched and explored from the perspective of consumer preferences (Luo et al, 2020). To analyze BB from a user preference perspective, this study presents a user preference mining algorithm that utilizes data mining (DM) and social behavior (SB). To explore users' personal preferences for the subject of BB, a cross-domain DM strategy was selected. This approach includes an examination of temporal behaviors to solve the issue of asynchrony in cross-domain behaviors. This paper is divided into the following four parts: a summary overview of the related research field, specifically personalized recommendation algorithms, and the current state of research in the BB-related field; implementation of the algorithms proposed in this study; validation of the efficacy of the methodology proposed in this study; and a summary overview of the entire study.

This investigation presents two new methodologies to enhance the understanding and prediction of user preferences. The first innovation involves using cross-domain user preference prediction based on users' SB information. This method goes beyond traditional domain-specific preference predictions, broadening the scope of analysis and enabling insights to be extracted from

diverse areas of user interaction. The second innovation integrates temporal behavior, recognizing that user preferences evolve over time and capturing their dynamic changes. Together, these innovations provide a comprehensive and dynamic perspective of user preferences, improving the effectiveness of predicting them and establishing a strong foundation for future research.

2. Related works

Personalized recommendation technology can be used to predict users' potential preferences by learning and mining their historical behaviors, which can help e-commerce websites make accurate product recommendations to improve product sales and the shopping experience. It can effectively help e-commerce websites achieve accurate product recommendations for users, thereby improving product sales and user shopping experience; moreover, it can be used in other fields to help improve the user experience. Wang et al. (Wang et al, 2022) proposed a potential preference recommendation method, which accurately mines potential preferences using users' behavioral similarity. Chen et al. (Chen et al, 2020) proposed a content-based personalized recommendation algorithm using collaborative filtering to improve the accuracy and effectiveness of recommendations in response to the problem of excessively homogeneous resource recommendations. Sardianos et al. (Sardianos et al, 2020) proposed a recommendation mechanism based on user preferences and habits to improve the trustworthiness of a recommender system, which is characterized by being explainable and persuasive, thereby increasing the recommender system's acceptance. Zou et al. (Zou et al, 2021) proposed a two-stage recommendation algorithm based on collaborative filtering and decomposition to solve the problem of poor diversity in the current recommendation system; the results of simulation experiments demonstrate that the algorithm is highly effective and accurate. Liu et al. (Liu et al, 2021) proposed a hybrid recommendation algorithm that combines collaborative filtering and genetic algorithms to achieve effective recommendations for target customers to mitigate the difficulty in combining manufacturing services. Gao et al. (Gao et al, 2020) proposed a time-series prediction method based on the K-nearest neighbor algorithm and support vector regression to predict the time of a user's subsequent purchase, which improves the effectiveness and accuracy by mining the temporal correlation of the user's consumption behavior. Nguyen et al.'s (Steenkamp et al, 2020) study on marketing behaviors and brand love revealed that investment in research and development, as well as investment in advertising, positively impacts brand love and that brand love can increase profitability and market value in the long term. Wu et al. (Wu et al, 2022) proposed a multi-context perception algorithm that integrates path and propagation methods to solve the problem of item representation in a knowledge graph in recommender systems. The algorithm efficiently learns user/item representations by introducing rules to describe user preferences and realizing high-order connectivity learning and local neighborhood feature expressions between users and items. Cai et al. (Cai et al, 2020) proposed a multi-objective hybrid recommendation model based on scores for the accuracy and diversity requirements of recommender systems, which simultaneously optimizes the accuracy, recall, diversity, novelty, and coverage. The experimental results revealed that the proposed algorithm provides more novel item recommendations by ensuring accuracy and diversity than the standard multi-objective optimization algorithm. Zhang (Zhang et al, 2021) proposed a Top-N recommendation algorithm integrated with a neural network to protect privacy from signal interference. The experimental results demonstrated the algorithm's potential value in privacy protection. Fang et al. (Fang et al, 2020) proposed a collaborative filtering recommendation algorithm based on deep neural network fusion to address the shortcomings of collaborative filtering

algorithms in terms of feature extraction and recommendation performance. They extracted text and other attributes using LSTM and deep neural networks, respectively, and integrated them into the feature matrix to improve the representation of users and objects. Their experiments demonstrated the excellent performance of this method.

The brand image often represents the brand's style, quality, level, and characteristics through the user's perspective, and is characterized by an abstract image to achieve the effect of visualization, which can effectively change the consumer's perception of the brand. Effective BB can have a beneficial effect in terms of marketing promotion. Steenkamp discussed the impact of the current rapid spread of Internet technology on global BB, presented recommendations and trends for global BB and management, discussed the key changes, and suggested future directions (Nguyen and Feng, 2020). Chatzopoulou et al.'s (Chatzopoulou and Navazhylava, 2022) exploratory study on branded digital media introduced a new concept of brand identity work, analyzed the relationship strategy between brand building and presentation, and provided suggestions for corporate brand presentation. On the contrary, Hemonnet-Goujot et al.t (Hemonnet-Goujot et al, 2022) analyzed the relationship between innovation and branding in the luxury industry based on consumer-brand and transgressive behavior theories and provided suggestions for sustainable innovation in luxury brands, which is more conducive to their brand building and development. Akhmedova et al. (Andersson et al, 2020) studied the brand effect in the sharing economy and proposed that brand trust positively contributes to the sharing economy. Andersson et al. addressed the issue of how relatively mature global organizations create value for their customers, exploring the role of BB therein and establishing their BB-based market orientation. Bose et al. (Bose et al, 2022) investigated the impact of regional identity on destination brand loyalty and investment attractiveness. They proposed defining the customer-based location brand asset dimension from the perspective of regional identity. The results demonstrated the significant role of BV in management and theory. Choi et al. (Choi and Seo, 2021) proposed a situation-based experimental design to analyze the changes in consumers' brand evaluations in the context of brand rumors in the restaurant industry and firms' response strategies. They found that rumors have a significant negative impact on brands, providing important academic and practical inspiration for the development of anti-rumor strategies. Qorbani et al. (Qorbani et al., 2021) proposed a new conceptual framework to analyze the relationship between the key elements of brand equity (BE) and customer equity (CE) and quantified the impact of marketing activities on CE. Through a questionnaire survey, they found that brand knowledge and differences positively affect customer acquisition and wallet share, and marketing activities significantly improve customer acquisition and wallet share through the mediating role of BE. Tran et al. (Tran et al, 2021) proposed a model of motivational sources influencing the use value of mobile applications in enterprise–consumer communication, thereby enhancing brand equity. Specifically, practical and hedonic motivations influence utility, which subsequently increases perceived quality, brand loyalty, brand awareness and associations.

In summary, facing the problem that existing recommendation algorithms focus too much on a single perspective, this study proposes a user preference mining algorithm based on DM and SB, which integrates the technical means of DM, deeply evaluates the user's SB, and obtains rich user preference information. In addition, this algorithm realizes cross-domain prediction based on user SB, which greatly expands the scope and depth of user preference prediction. This comprehensive and in-depth analysis increases the accuracy and practicality of user preference prediction. In addition, this prediction method provides a new perspective for understanding user behavior and

strong data support for brand building. Brands can better understand the target user group based on user preference prediction, further optimize product design and marketing strategies, and enhance BV. With its unique perspective and strong practicality, this user preference algorithm based on DM and SB will undoubtedly provide new ideas and methods for future recommendation systems and brand building.

3. BB for preference mining algorithms based on DM and SB

In this study, a cross-domain user-preference prediction algorithm is proposed based on social and e-commerce platform data. For the massive data problem frequently faced in the cross-domain, first, a classification method based on the Hamming distance is used for screening, and then, a cross-domain user preference prediction model is built considering the timing problem to predict the user-brand preference—to provide data support for the enterprise's BB.

3.1. DM methods based on user classification

To achieve accurate prediction of user preferences in the face of massive data, it is necessary to effectively screen and acquire useful data while discarding useless data. This allows for the construction of personalized user profiles that include behavioral habits, personal preferences, and other relevant indexes. By doing so, accurate one-to-one prediction of user preferences can be achieved. In this study, the user–user relationship is explored first, followed by the exploration of user–commodity and commodity–commodity relationships; then, user preferences are analyzed from the perspective of data, and the neural network method is adopted to establish a user group preference value model to accurately predict user preferences. In this study, data are first coded. The distance between codes is frequently referred to as the Hamming distance, which can be used to compare similarities and differences between codes and coded information. Overall, the greater the Hamming distance, the greater the difference between codes. Therefore, to study the user's preference from the perspective of data, several indicators are assigned to the user's personal behavior; the encoding of the user's behavior that occurs is set to 1, and the encoding of the user's behavior that does not occur is set to 0. The preliminary analysis of the data reveals that the "media behavior" and "video behavior" in the user's data can be used as an indicator of the user's personal behavior to a certain extent. Additionally, it reveals that the "media behavior" and "video behavior" in the user data can indicate the 'user's personal preference to a certain extent, while the "e-commerce "behavior" can reflect the 'user's brand preference; hence, this study performs the coding of personal preference indexes for the three kinds of behaviors, and the coding rules are presented in Fig. 1.

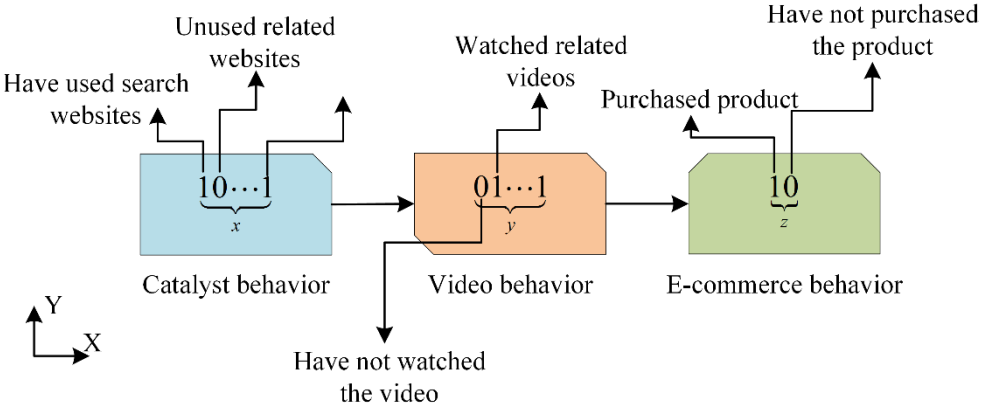


Fig. 1 Hamming distance coding rules.

By encoding the user's behavioral data, the similarity and difference of preference between users can be judged by the size of the Hamming distance between them, which is equal to the absolute value after subtracting the code element and its counterpart. Then, by adding them up, the larger the Hamming distance, the greater the difference in preference between the two users. The size of the Hamming distance between the encoded user and user can be set as a threshold, and then, by determining whether the difference between them belongs to the same threshold, we can determine whether the two users belong to the same category. In the process of calculating the Hamming distance, two elements are first selected from the set U, which are defined as the two most similar elements. Their initial set is presented in Eq. (1).

$$C = \{u_i, u_j\} \quad (1)$$

In Eq. (1), u_i and u_j are the two most similar elements, and C is the initial set of this definition. Then, any element u_k is selected from the remaining variables when the selected u_k satisfies the condition presented in Eq. (2).

$$\{|u_i - u_j| < \&, |u_j - u_k| < \&, \cdots\} \quad (2)$$

In Eq. (2), $\&$ denotes the Hamming distance threshold set in this study. If the u_k selected at this point satisfies the conditions of Eq. (2), the u_k is grouped into set, as presented in Eq. (3).

$$C = \{u_i, u_j, u_k\} \quad (3)$$

After traversing all the elements in the set, all the elements in the resulting set C are classified into one category. The most important aspect of the user classification process is the selection of threshold values—achieved by selecting different thresholds at which various users' behavioral data can be clearly classified. Fig. 2 illustrates a schematic diagram of the sequence numbers of the target user groups under this classification rule.

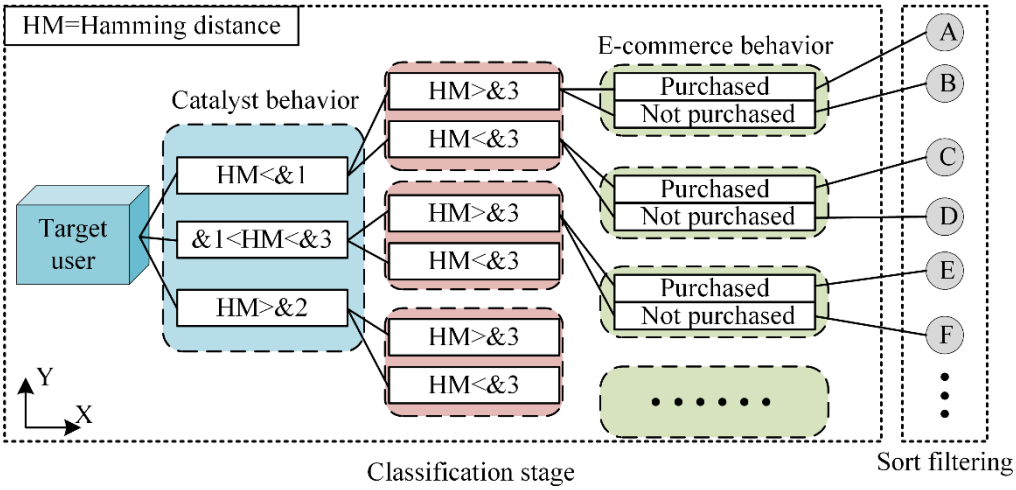


Fig. 2 Classification rule diagram.

After classifying the data, the data are used as a training set. After training, the BP neural

network is used to predict the test samples. The eigenvalues of the users in the training set are used as 1×146 vectors, which are finally constructed as 146×4790 matrices and used as input data, and the numbers corresponding to the target user groups constructed in the previous section are used as output data to build the network. Each user's data are analyzed, their eigenvector values are obtained, the threshold value is set according to the different eigenvector values of their corresponding categories, the required data are filtered out, and useless data are eliminated. Fig. 3 presents a schematic diagram of the BP network.

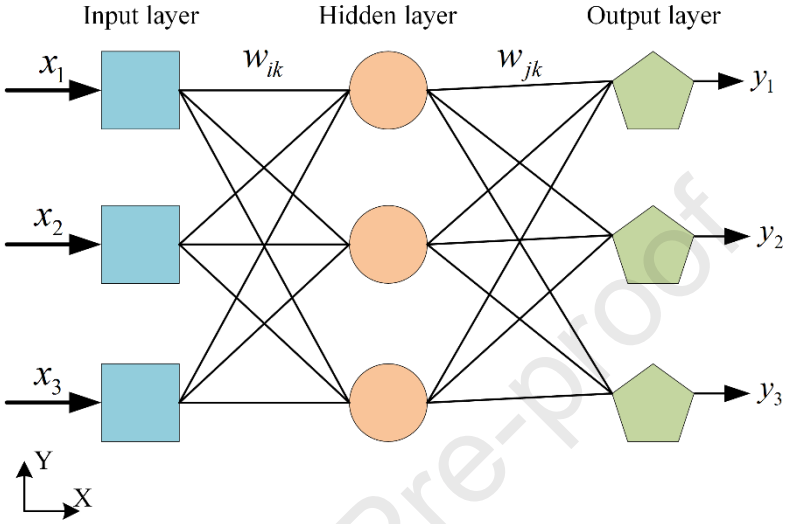


Fig. 3 BP neural network schematic diagram.

The neural network training process involves analyzing each user's data to obtain their eigenvector values. Then, by setting thresholds, users are matched to categories based on differences in their feature vector values. This method filters out the required data and eliminates irrelevant or useless data. The backpropagation neural network uses a supervised learning method to adjust the weight and bias of the network, aiming to make the predicted output as close as possible to the actual output. The network updates the weight and bias during the training process based on the error gradient, gradually improving the prediction accuracy using the backpropagation algorithm.

3.2. Preference mining algorithm based on DM as well as SB and BB

Typically, the prediction of user preferences is limited to a single domain, such as the prediction of user film preferences based on their viewing habits. Currently, large e-commerce platforms, including Amazon and eBay, predict user preferences more accurately by combining data from multiple cross-domains. Cross-domain recommendation is often used to help optimize the data scarcity and cold-start problem caused by the scarcity of information in the target domain, and this method plays an important role in the initial construction of new users, systems, or products (Zhang et al, 2021). The data distribution in the cross-domain prediction method can be divided into categories based on whether users and projects across different domains overlap, assuming that two domains exist, wherein the user sets are U_s and U_t, and the project sets are I_s and I_t. Fig. 4 presents the results.

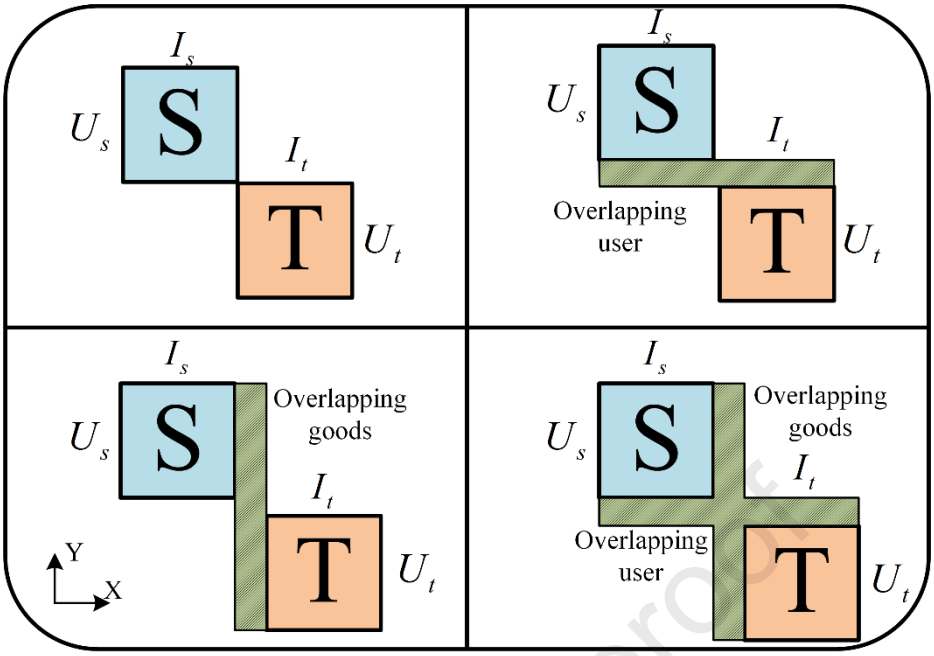


Fig. 4 Cross-domain task data distribution diagram.

To implement cross-domain information prediction, the amount of data is usually large owing to the excessive number of domains. Therefore, the massive amount of data is filtered using the classification method mentioned in the previous section to facilitate subsequent processing. To tackle the heterogeneity of cross-domain data, we propose a cross-domain time-series algorithm for predicting user preferences based on their social domain information. Additionally, we introduce a study of time-series behaviors to address the issue of asynchronous behaviors across different domains. In this study, an improved cross-domain preference ranking model based on factorization machines (FM) is proposed (Guo et al, 2022). Fig. 5 presents the framework flow of the improved cross-domain personalized preference prediction (CDPPP) model.

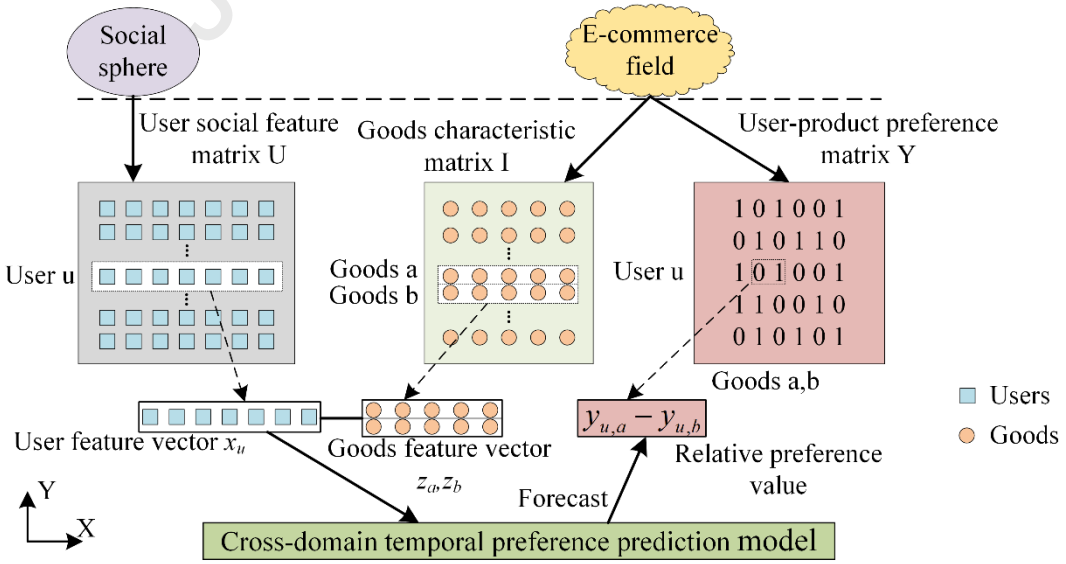


Fig. 5. Frame diagram of cross-domain personalized preference prediction model.

Based on the BB topic investigated herein, the developed model does not predict the user's

preference for each brand but, instead, for the user's different degrees of preference for the brand in this study, as well as for another brand. Ultimately, it establishes a computational process for predicting the user's preference for the brand, as Eq. (4) indicates.

$$P_{u,a\succ b}=\frac{1}{2}\left(y_{u,a\succ b}+1\right) \quad (4)$$

In Eq. (4), A denotes the relative probability that the user prefers brand a more than brand b, $y_{u,a\succ b}$ denotes the relative preference degree of the user, and $a\succ b$ denotes that the user has a significantly greater preference for a than for b. Further, the process of calculating the relative interest preference for the two different brands is presented in Eq. (5).

$$\hat{y}_{u,a\succ b}(x,z)=\sum_{i=1}^{m}u_i x_i+\sum_{i=1}^{n}u_i^{'}z_i+ \frac{1}{2}\sum_{k=1}^{l}\left(\left(\sum_{i=1}^{m}v_{i,k}x_i+\sum_{i=1}^{n}v_{i,k}^{'}z_i\right)^2-\sum_{i=1}^{m}v_{i,k}^2x_i^2-\sum_{i=1}^{n}v_{i,k}^{'2}z_i^2\right) \quad (5)$$

In Eq. (5), x_u is the feature vector of a certain user u, $z=z_a-z_b$ is the difference vector of the two commodities, z_a and z_b are the vectors of the two branded commodities, $\hat{y}_{u,a\succ b}(x,z)$ is the relative preference of the user for the two branded commodities predicted by the model, and $u,u^{'},V,V^{'}$ is the parameter to be trained in the model. The Sigmoid regression function is used to transform the prediction results, transforming their form from a numerical value to the probability that the user prefers brand a over brand b, as Eq. (6) indicates.

$$\hat{p}_{u,a\succ b}=\sigma\left(\hat{y}_{u,a\succ a}\right)=\frac{1}{1+e^{-\hat{y}_{u,a\succ b}}} \quad (6)$$

In Eq. (6), $\hat{p}_{u,a\succ b}$ is the probability that a user prefers brand a over brand b. Then, the cross-entropy loss function (Cross-Entropy, CE) is used to transform the error between the predicted and actual values into a loss value that can be used by the model. The details are presented in Eq. (7).

$$-p_{u,a\succ b}\log\left(\hat{p}_{u,a\succ b}\right)-\left(1-p_{u,a\succ b}\right)\log\left(1-\hat{p}_{u,a\succ b}\right) \quad (7)$$

Finally, the selection method for stochastic gradient descent (SGD) is used for the gradient in the direction of the error, and the results are used to update the model parameters with the updated equations for all parameters. Eq. (8) indicates the expression.

$$\begin{cases} u_i \leftarrow u_i - \eta\left(\sigma\left(\hat{y}_{u,a\succ b}\right) - p_{u,a\succ b}\right)\cdot x_i \\ u_i^{'} \leftarrow u_i^{'} - \eta\left(\sigma\left(\hat{y}_{u,a\succ b}\right) - p_{u,a\succ b}\right)\cdot z_i \\ v_{i,k} \leftarrow v_{i,k} - \eta\left(\sigma\left(\hat{y}_{u,a\succ b}\right) - p_{u,a\succ b}\right)\left(x_i \sum_{j=1}^{m} v_{j,k} x_j - v_{i,k} x_i^2 \right) \\ v_{i,k}^{'} \leftarrow v_{i,k}^{'} - \eta\left(\sigma\left(\hat{y}_{u,a\succ b}\right) - p_{u,a\succ b}\right)\left(z_i \sum_{j=1}^{m} v_{i,k}^{'} z_j - v_{i,k}^{'} z_i^2 \right) \end{cases} \quad (8)$$

The user's SB data is processed by considering temporal features. To ensure objectivity and scientificity, demographic, social temporal, and user dimensional features are selected to construct the user's temporal feature vector, as shown in Eq. (9).

$$x_u^h = \left(a_u, d_u^h, c_u\right) = \left(\underbrace{1,\cdots,0.2,\cdots,0.5}_{a_u}, \underbrace{-0.7,\cdots,0.5,-0.1}_{d_u^h}, \underbrace{0,\cdots,0,\cdots,1}_{c_u} \right) \quad (9)$$

In Eq. (9), x_u^h denotes the user temporal feature vector of user u, a_u denotes the demographic features, d_u^h denotes social temporal features, and c_u denotes user dimensional features. The commodity's feature vector is more stable than its user features; therefore, in constructing the former, chronological factors' influence can be disregarded based on the principles of objectivity and scientificity, coupled with the consideration of the BV in this study, final choice of the commodity category features, shopping context features, and brand dimension features, to construct the commodity feature vectors, as presented in Eq. (10) (Khamis et al, 2020).

$$z_i = \left(a_i, d_i, c_i\right) = \left(\underbrace{0,\cdots,1,\cdots,0}_{a_i}, \underbrace{0.4,\cdots,-0.6,0.2}_{d_i}, \underbrace{0,\cdots,0,\cdots,1}_{c_i} \right) \quad (10)$$

。In Eq. (10), z_i denotes the product feature vector, a_i denotes the product category feature, d_i denotes the shopping context feature, and c_i denotes the brand dimension feature. As the user's interest is not static and often changes over time, considering the temporal sequence of user-brand preference features is also necessary. Finally, based on whether the user has purchased the brand, the user's time-series preference feature vector for brand is established, as indicated in Eq. (11).

$$y_u^h = \left(0,0,1,\cdots,0,1,1\right) \quad (11)$$

In Eq. (11), y_u^h denotes the user's temporal preference vector for the brand; 1 denotes the brand that the user has purchased, and 0 denotes the brand that the user has not purchased. Finally, the model is trained based on the user temporal feature vector x_u^h, product feature vector z_i, and user temporal preference vector for brands y_u^h established in this study; the final mapping

relationship reached by the model for user preference is $f\left(x_u^h, z_i\right) \to y_{u,i}^h$. Finally, based on the establishment of temporal feature vectors, this study proposes the cross-domain temporal preference prediction (CDTPP) model. Fig. 6 presents the CDTPP model's framework and its flow.

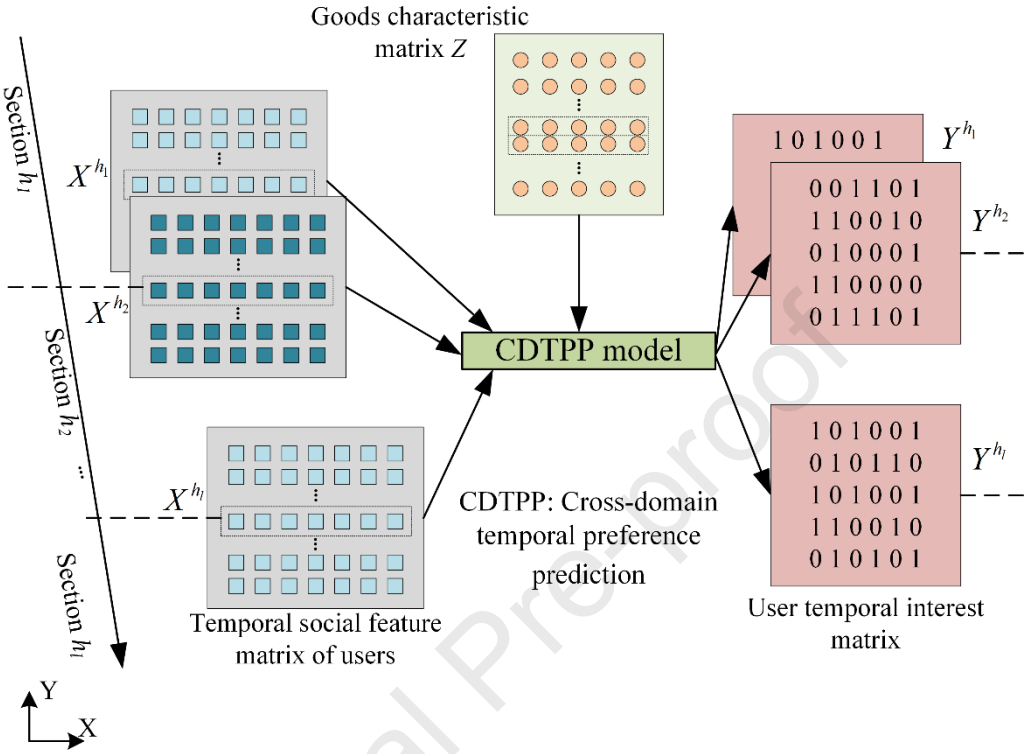


Fig. 6. Schematic diagram of cross-domain temporal preference prediction model.

Setting the user's temporal feature vector as x_u^h and selecting two brands of goods as a and b, respectively, these two goods' corresponding feature vectors are z_a and z_b. The calculation of the relative preference for the two brands at this time is presented in Eq. (12).

$$\hat{y}_{u,a\succ b}^h\left(x^h, z\right) = \sum_{i=1}^{m} u_i x_i^h + \sum_{i=1}^{n} u_i^{'} z_i + \frac{1}{2}\sum_{k=1}^{l}\left(\left(\sum_{i=1}^{m} v_{i,k} x_i^h + \sum_{i=1}^{n} v_{i,k}^{'} z_i\right)^2 - \sum_{i=1}^{m} v_{i,k}^2 x_i^{h2} - \sum_{i=1}^{n} v_{i,k}^{'2} z_i^2\right) \quad (12)$$

In Eq. (12), $z = z_a - z_b$ denotes the difference between the two product vectors, $\hat{y}_{u,a\succ b}^h\left(x, z\right)$ denotes the output brand relative preference level, and $u, u^{'}, V, V^{'}$ denotes the parameters to be trained in the model. As mentioned previously, a Sigmoid function is used to transform the representation from numerical values to probabilities, as presented in Eq. (13).

$$\hat{p}_{u,a\succ b}^h = \sigma\left(\hat{y}_{u,a\succ b}^h\right) = \frac{1}{1+e^{-\hat{y}_{u,a\succ b}^h}} \quad (13)$$

The same applies the cross-entropy loss function, which addresses the gap between the predicted and actual values and transforms them into loss values that can be used by the model, as presented in Eq. (14).

$$-p_{u,a\succ b}^{h}\log\left(\hat{p}_{u,a\succ b}^{h}\right)-\left(1-p_{u,a\succ b}^{h}\right)\log\left(1-\hat{p}_{u,a\succ b}^{h}\right) \quad (14)$$

In Eq. (14), $p_{u,a\succ b}^{h}$ represents the relative preference probability of the user for the two brands, and $p_{u,a\succ b}^{h}=\frac{1}{2}\left(y_{u,a\succ b}^{h}+1\right)$. Finally, the stochastic gradient descent method is used to obtain the gradient in the direction of the error, which is used to update the model parameters—calculated as presented in Eq. (15).

$$\begin{cases} u_i \leftarrow u_i-\eta\left(\sigma\left(\hat{y}_{u,a\succ b}^{h}\right)-p_{u,a\succ b}^{h}\right)\cdot x_i^{h} \\ u_i^{'} \leftarrow u_i^{'}-\eta\left(\sigma\left(\hat{y}_{u,a\succ b}^{h}\right)-p_{u,a\succ b}^{h}\right)\cdot z_i \\ v_{i,k} \leftarrow v_{i,k}-\eta\left(\sigma\left(\hat{y}_{u,a\succ b}^{h}\right)-p_{u,a\succ b}^{h}\right)\left(x_i^{h}\sum_{j=1}^{m}v_{j,k}x_j^{h}-v_{i,k}x_i^{h2}\right) \\ v_{i,k}^{'} \leftarrow v_{i,k}^{'}-\eta\left(\sigma\left(\hat{y}_{u,a\succ b}^{h}\right)-p_{u,a\succ b}^{h}\right)\left(z_i\sum_{j=1}^{m}v_{i,k}^{'}z_j-v_{i,k}^{'}z_i^{2}\right) \end{cases} \quad (15)$$

Based on the cross-domain temporal preference prediction model established above, this study classifies and rounds off the user's historical behavioral data and, thereafter, randomly selects user-commodity data pairs within a random interval, which is used as a training set to train the model. After the training is completed, the social information and e-commerce platform information are collected from overlapping user $u\in U_1\cup U_2$ on an e-commerce platform and a social platform, and a user temporal feature vector x_u^{h} is constructed, which can be used to predict the user's preference for the current brand y_u^{h}. Based on the cross-validation method, the model's optimal parameter settings are selected. Table 1 presents the specific parameter settings.

Table. 1 Model parameter setting

BP neural network		Factorization machines	
Parameter	Detail	Parameter	Detail
Number of hidden layer nodes	15	The length of the factor vector	37
Learning rate	0.01	Regularization parameter	0.05
Epoch	150	Learning rate	0.01

4. Evaluation of DM- and SB-based user preference mining algorithms in BB

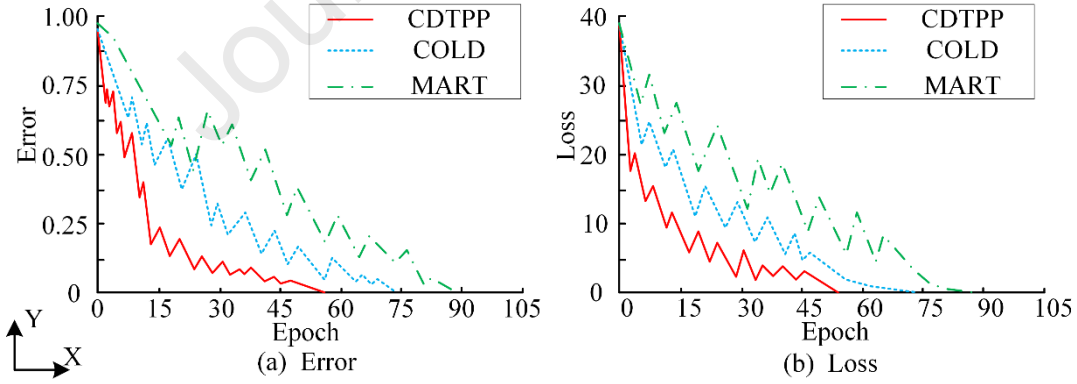
To further validate the method proposed in this study, the data of users logging in with social accounts on a large domestic e-commerce platform are selected, containing the purchase data of 5,234 different brands of goods in this batch of users, wherein the ID of each user, ID of the

corresponding brand, and purchase time data are recorded. Through the open-source API interface of the social media platform, all the records of this group of users are obtained, including the text content in their sends, retweets, and comments. The data is screened using the method proposed in this study. Data with excessively large time differences is eliminated, and only the part with overlapping time for the experiment is selected. Users with less than 20 texts on social platforms and less than 10 purchase records on e-commerce platforms are eliminated to avoid the influence of the data on the model. The users' social platform texts are processed to filter out relatively abstract terms with special symbols in the network and all deactivated words. Table 2 presents the final user dataset. Notably, 80% of the dataset is randomly selected as the training set, and the remaining 20% is selected as the test set.

Table. 2 Dataset details.

Number of overlapping users	Brand type	Details		
		Platform	Type	Amount (million)
7,851	3,418	Social platform	Text information	2.24
		E-commerce platform	Purchase record	1.17

The computing power cost-aware online and lightweight deep pre-ranking system (COLD) (Kang et al, 2021) and multiple additive regression tree (MART) (Slade and Naylor, 2020) models are selected for comparison with the CDTPP model proposed in this study, and the parameter settings of the model are derived from the optimal settings given in the literature. The MART model is compared with the CDTPP model proposed in this study. A comparison is first performed for the convergence of the algorithms, the results of which are presented in Fig. 7, which indicates that the algorithm proposed in this study can reach the optimal state after 57 training iterations—reduced by 14 and 28 compared with the COLD and MART models, respectively.



Note: COLD means The computing power cost-aware Online and Lightweight Deep pre-ranking system, MART represents Multiple Additive Regression Tree, and CDTPP represents the cross-domain temporal preference prediction model proposed by the institute.

Fig. 7 Convergence of the three algorithms. (a) error, and (b) loss.

A dataset from another e-commerce platform is selected and used to test the untrained model's out-of-sample accuracy. The model's convergence value mean and standard deviation are obtained and presented in Table 3. The results show that the proposed algorithm's convergence is superior to that of the other two models. The model proposed demonstrates higher training efficiency and can achieve better results at a lower cost.

Table. 3 Mean and standard deviation of convergence of the three algorithms

Number of experiments		CDTPP	COLD	MART
1	Mean	**4.15E-06**	7.94E-05	4.26E-04
	Standard	**5.91E-05**	5.16E-04	2.38E-03
2	Mean	**7.62E-06**	5.91E-05	8.24E-04
	Standard	**8.69E-04**	6.46E-04	3.05E-03
3	Mean	**5.97E-06**	4.39E-04	6.95E-03
	Standard	**1.95E-04**	5.79E-03	2.69E-02
4	Mean	**2.49E-05**	6.32E-04	5.71E-03
	Standard	**7.13E-05**	5.47E-04	8.09E-03
5	Mean	**5.61E-04**	4.90E-03	7.52E-02
	Standard	**8.73E-05**	5.37E-03	2.72E-02

Note: Mean indicates The mean of convergence performance, Standard indicates the standard deviation of convergence performance, COLD indicates the computing power cost-aware Online and Lightweight Deep pre-ranking system, MART represents Multiple Additive Regression Tree, and CDTPP represents the cross-domain temporal preference prediction model proposed by the institute.

Subsequently, the mean square error (MSE) and mean absolute error (MAE) of the three algorithms are compared, the results of which are presented in Fig. 8. This table describes an ordered classification problem, where categories have a natural order, such as ratings from 0 to 5. The MAE can be used to measure the absolute deviation between predicted and actual classes, while the MSE can impose greater penalties for larger errors. Fig. 8 reveals that the three algorithms' MSE and MAE values decrease with an increase in the number of iterations; however, the proposed algorithm exhibits a more evident decreasing trend and a lower minimum value, which indicates that it has a lower error and a higher degree of goodness-of-fit in practical applications.

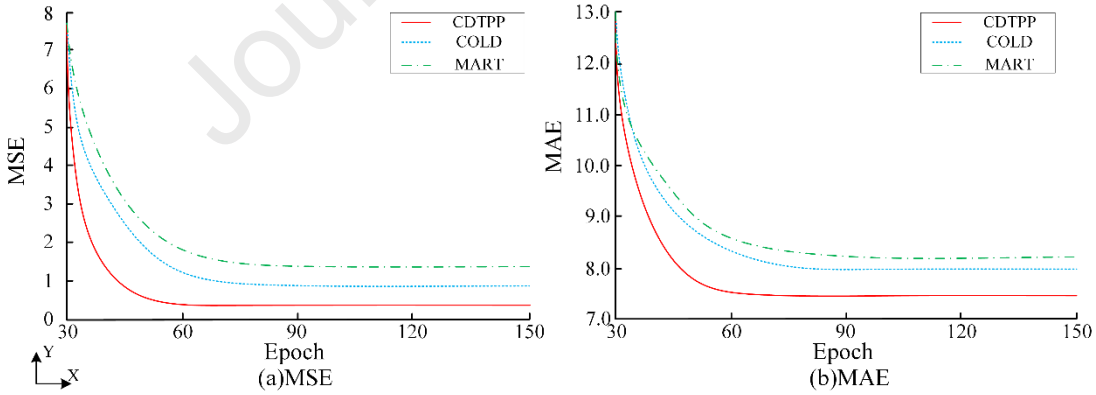


Fig. 8 (a) MSE & (b) MAE of the three algorithms.

The three algorithms' AUC values and accuracy rates are tested; to reduce the impact of errors, the three algorithms are tested five times and the average value is considered for comparison. The results—presented in Table 4—indicate that the average AUC value of the proposed CDTPP model reaches 0.953, which is 0.027 and 0.044 higher than those of the COLD and MART models, respectively, and the average accuracy rate is 0.984, which is 0.016 and 0.056 higher than those of

the COLD and MART models, respectively.

Table. 4 AUC mean and precision mean of three algorithms

Index	Number of experiments	Model		
		CDTPP	COLD	MART
AUC	1	0.949	0.927	0.902
	2	0.953	0.934	0.906
	3	0.951	0.919	0.911
	4	0.956	0.928	0.916
	5	0.958	0.926	0.914
	Average value	**0.953**	**0.926**	**0.909**
Accuracy	1	0.981	0.964	0.921
	2	0.997	0.968	0.926
	3	0.982	0.971	0.938
	4	0.985	0.973	0.92
	5	0.979	0.965	0.938
	Average value	**0.984**	**0.968**	**0.928**

Note: AUC means area under curve, COLD means The computing power cost-aware Online and Lightweight Deep pre-ranking system, MART represents Multiple Additive Regression Tree, and CDTPP represents the cross-domain temporal preference prediction model proposed by the institute.

Thereafter, the three algorithms' fits are tested, the results of which are presented in Fig. 9, which indicates that the fit of the proposed model—with a fit value of 98.87%—is 4.12% and 11.26% higher and, thus, better than those of the COLD and MART models, respectively. This suggests that the proposed model's prediction results for user preferences match the actual situation well and have better practical significance in actual use than those of the other models.

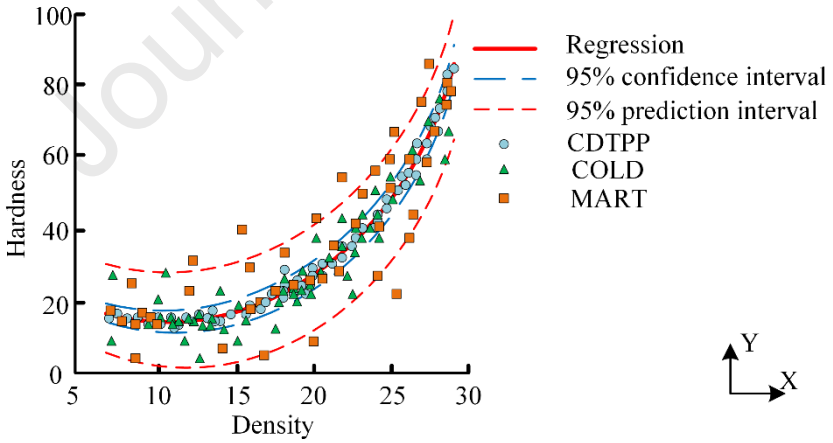


Fig. 9. Fit degree of the three algorithms.

Finally, simulation experiments are conducted to test the accuracy of the proposed model's prediction of the degree of user preference for brands, and to test the model's out-of-sample performance. The range of the user preference index for the brand is set as 0–5, which indicates that the user ranges from having no interest in the brand to being highly accustomed to it; accordingly, 15 experiments are conducted to obtain each model's error value, the results of which are presented in Table 5, which reveals that the CDTPP model proposed herein is highly effective in predicting

the user's preference index for the brand, with an error of only 0.11—0.20 and 0.42 higher than that of the COLD and MART models, respectively. This suggests that the algorithms proposed herein exhibit an extremely small error in practical application, and the output results are relatively more accurate.

Table 5 Comparison of prediction error results of preference index

Sample number	CDTPP	COLD	MART	Actual preference index
1	4.8	4.5	4.1	4.9
2	4.4	4.8	3.8	4.5
2	3.6	3.5	3.1	3.7
4	3.8	3.5	3.4	3.9
5	1.5	1.9	2.1	1.4
6	0.8	1.1	1.4	0.7
7	4.7	4.5	4.1	4.8
8	4.6	4.1	3.9	4.5
9	3.9	4.0	4.3	3.7
10	3.7	4.1	4.6	3.5
11	4.1	4.5	3.8	4.2
12	2.8	2.5	2.9	2.7
13	2.5	2.3	2.5	2.4
14	2.9	2.7	2.9	2.8
15	2.2	2.0	1.9	2.1
Average error	**0.11**	**0.31**	**0.53**	**-**

Note: COLD means The computing power cost-aware Online and Lightweight Deep pre-ranking system, MART represents Multiple Additive Regression Tree, and CDTPP represents the cross-domain temporal preference prediction model proposed by the institute.

Finally, for the CDTPP model proposed herein, the personnel of a brand party is invited to evaluate its actual role in BB, and 10 of its managers, 15 of its employees, and 10 of its experts in the relevant fields are invited to evaluate it, the results of which are presented in Fig. 10. During the evaluation process, participants assess the model's utility. Management evaluates the model's impact on BD for strategic decision-making, frontline employees evaluate its role in improving work efficiency and productivity, and industry experts assess its innovation and potential for improvement in light of industry trends and challenges. A rating scale ranging from 0 (lowest) to 100 (highest) is used to indicate the model's effectiveness. The evaluation results of all participants are comprehensively analyzed to fully understand the model's usefulness. This scientific and practical evaluation method helps us to identify the model's strengths and weaknesses, allowing us to continuously optimize and improve its practicality in brand building. As Fig. 10 indicates, the ratings of the CDTPP model proposed herein are high across all three categories of personnel who evaluated it. Specifically, high score levels are observed among managers and experts in related fields. In summary, the CDTPP model proposed herein can more accurately predict the degree of user preference for a particular brand, which can provide more accurate data support for the BB of an enterprise and positively contribute to its BB.

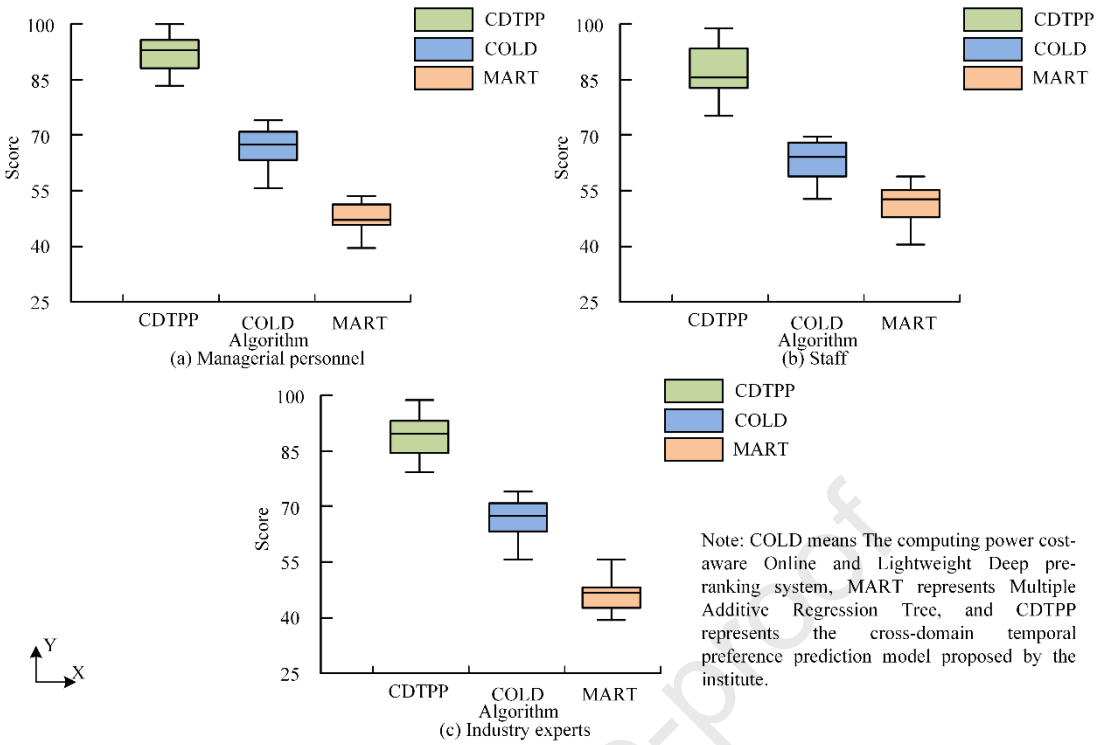


Fig. 10 Evaluation of the three algorithms. (a) managerial personnel, (b) staff, and (c) industry experts.

5. Managerial implications

User preference mining algorithms, when combined with DM and SB analysis, play a significant role in this process. Accurately grasping and utilizing insights into consumer preferences is crucial in BD management and practice. It is important to maintain a clear and logical structure, avoiding sprawling descriptions and complex terminology. By analyzing social media data, enterprises can discover user behavioral patterns, emotional attitudes, and consumption trends. This information is essential for precise customer targeting and refined market segmentation. A multidimensional analysis of user preferences lays the foundation for personalized marketing and provides data support for product and service iterations. This allows brands to modify product features according to customer feedback, improving the overall customer experience.

In addition, SB analysis offers a new approach to predicting market trends and enhancing a brand's ability to understand market dynamics, thereby maintaining a competitive advantage in a highly competitive environment. With regard to risk management, real-time monitoring and analysis of SB data helps brands identify and respond to potential negative publicity early, thus reducing the risk of brand crises. From a sales and conversion perspective, understanding user preferences helps brands design more effective marketing campaigns, improve conversion rates, and ultimately increase sales. Furthermore, incorporating user preference data into customer relationship management systems promotes interaction between the brand and its customers, leading to enhanced customer satisfaction and loyalty.

In summary, using user preference mining algorithms based on DM and SB in brand building not only provides brands with a comprehensive platform for understanding and serving consumers, but also offers scientific data support for the brand's sustained growth and success in the market.

6. Conclusion

This study suggests a user preference mining method based on DM and SB for predicting user preferences, with the objective of providing data support for BB of businesses to solve the absence of BV construction in contemporary SMEs. The experimental results demonstrated that compared with the COLD and MART models, the proposed algorithm has better training efficiency and convergence; a lower mean square error and average absolute error, indicating that its output results will be more accurate; an AUC value of 0.953, indicating that its prediction is extremely real; an accuracy rate of 0.928; and a fit of 98.87%, indicating that it is more suitable for predicting consumer preferences. The model proposed herein exhibited high scores across several parameters, allowing industry stakeholders to estimate the value of the algorithm for enterprise BB. In the simulation studies, the prediction of the user's brand preference was accurate, with an average error of only 0.11. Although this study has made valuable contributions, it also has some limitations. For instance, due to the complexity and diversity of the data, the algorithm may not be able to make accurate predictions for all types of user data. Additionally, the algorithm may face limitations in computing resources when processing large-scale data. In the future, the use of DM and user preference mining algorithms for SB in brand building is expected to become more widespread. With further optimization of the algorithm and improvements in computing resources, it is expected that this algorithm will be applied in several other fields, providing more powerful data support for brand building. Simultaneously, as user data continues to grow, extracting valuable information from this data will become an important direction for future research.

References

Akhmedova, A., Vila-Brunet, N., Mas-Machuca, M., et al, 2021. Building trust in sharing economy platforms: trust antecedents and their configurations. INTERNET RES, 31 (4), 1463-1490.

Andersson, S., Awuah, G.B., Aagerup, U., et al, 2020. How do mature born globals create customer value to achieve international growth? INT MARKET REV., 37 (2), 185-211.

Bose, S., Pradhan, S., Bashir, M., et al, 2022. Customer-Based Place Brand Equity and Tourism: A Regional Identity Perspective. J TRAVEL RES., 61 (3), 511-527.

Cai, X., Hu, Z., Chen, J., 2020. A many-objective optimization recommendation algorithm based on knowledge mining. INFORM SCIENCES., 537 (1), 148-161.

Chatzopoulou, E., Navazhylava, K., 2022. Ethnic brand identity work: Responding to authenticity tensions through celebrity endorsement in brand digital self-presentation. J BUS RES., 142 (3), 984-987.

Chen, S., Huang, L., Lei, Z., et al, 2020. Research on personalized recommendation hybrid algorithm for interactive experience equipment. COMPUT INTELL-US., 36 (3), 1348-1373.

Choi, J., Seo, S., 2021. Do brand rumors matter? The role of brand equity and response strategy to brand rumor. INT J CONTEMP HOSP M., 33 (8), 2862-2879.

Fang, J., Li, B., Gao, M., 2020. Collaborative filtering recommendation algorithm based on deep neural network fusion. INT J SENS NETW., 34 (2), 71-80.

Findley, F., Johnson, K., Crang, D., et al, 2020. Effectiveness and Efficiency of TV's Brand-Building Power: A Historical Review: Why the Persuasion Rating Point (PRP) Is a More Accurate Metric than the GRP. JAR., 60 (4), 361-369.

Gao, H., Kuang, L., Yin, Y., et al, 2020. Mining consuming Behaviors with Temporal Evolution for Personalized Recommendation in Mobile Marketing Apps. MONET., 25 (4), 1233-1248.

Guan, J., Wang, W., Guo, Z., et al, 2021. Customer experience and brand loyalty in the full-service hotel sector: the role of brand affect Customer experience and brand loyalty. INT J CONTEMP HOSP M。, 33 (5), 1620-1645.

Guo, Y., Mustafaoglu, Z., Koundal, D., 2022. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. JCCE., 2 (1), 5－9.

He, J., Calder, B.J., 2020. The experimental evaluation of brand strength and brand value. J BUS RES., 115 (7), 194-202.

Hemonnet-Goujot, A., Kessous, A., Magnoni, F., 2022. The effect of sustainable product innovation on the consumer－luxury brand relationship: The role of past identity salience.J BUS RES., 139 (2), 1513-1524.

Kang, S.K., Hwang, J., Kweon, W., et al, 2021. Item-side Ranking Regularized Distillation for Recommender System. INFORM SCIENCES., 580 (1), 15-34.

Khamis, M.A., Ngo, H.Q., Nguyen, X., et al, 2020. Learning Models over Relational Data Using Sparse Tensors and Functional Dependencies. ACM T DATABASE SYST., 121 (1), 161-168.

Kolbl, I., Diamantopoulos, A., Arslanagic-Kalajdzic, M., et al, 2020. Do brand warmth and brand competence add value to consumers? A stereotyping perspective. J BUS RES., 118 (9), 346-362.

Liu, Z., Wang, L., Li, X., et al, 2021. A multi-attribute personalized recommendation method for manufacturing service composition with combining collaborative filtering and genetic algorithm. J MANUF SYST., 58 (1), 348-364.

Luo, J., Dey, B.L., Yalkin, C., et al, 2020. Millennial Chinese consumers' perceived destination brand value. J BUS RES., 116 (8), 655-665.

Nguyen, H.T., Feng, H., 2021. Antecedents and financial impacts of building brand love. IJRM., 38 (3), 572-592.

Qorbani, Z., Koosha, H., Bagheri, M., 2021. An integrated model for customer equity estimation based on brand equity. INT J MARKET RES., 63 (5), 635-664.

Sardianos, C., Varlamis, I., Chronis, C., et al, 2020. The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency. INT J INTELL SYST., 36 (2), 656-680.

Slade, E., Naylor, M.G., 2020. A fair comparison of tree－based and parametric methods in multiple imputation by chained equations. SIM MEDLINE., 39 (8), 1156-1166.

Steenkamp, J.B.E.M., 2020. Global Brand Building and Management in the Digital Age. J INT MARKETING, 28 (1), 13-27.

Tran, T.P., Mai, E.S., Taylor, E.C., 2021. Enhancing brand equity of branded mobile apps via motivations: A service-dominant logic perspective. J BUS RES, 125 (2), 239-251.

Wang, C., Yang, Y., Suo, K., et al, 2022. MulSetRank: Multiple set ranking for personalized recommendation from implicit feedback. KNOWL-BASED SYST., 249 (5), 1-11.

Wu, C., Liu, S., Zeng, Z., et al, 2022. Knowledge graph-based multi-context-aware recommendation algorithm. INFORM SCIENCES., 595 (1), 179-194.

Zhang, H., Dong, Y., Li, J., et al., 2021. An efficient method for time series similarity search using binary code representation and hamming distance. IDA., 25 (2), 439-461.

Zhang, L., 2021. Top-N recommendation algorithm integrated neural network. NCA., 33 (9), 3881-3889.

Zou, F., Chen, D., Xu, Q., et al, 2021. A two-stage personalized recommendation based on multi-objective teaching – learning-based optimization with decomposition. NEUROCOMPUTING., 452 (10), 716-727.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: