Hindawi

Research Article

Prediction Method of User Behavior Label Based on the BP Neural Network

Ruihang Shen (ID)

Northwestern University, Evanston, USA

Correspondence should be addressed to Ruihang Shen; ruihangshen2019@u.northwestern.edu

With the development of big data, precision marketing helps businesses to be more efficient in selling products and stand out from the fierce competition. In this paper, the backpropagation neural network is introduced as an approach to analyze the data of user online behaviors and create labels for each user. In this way, the business will be able to realize accurate user classification and forecast users' future behaviors and thus achieve precision marketing. The study tested the backpropagation neural network with a real user behavior dataset from Taobao for a recommendation. According to users' behavior data, the network successfully classified users into 5 clusters with distinct labels and this information can give the business valuable insights into their customers for precision marketing and selling products.

1. Introduction

In the nineteenth century, a successful businessman, John Wanamaker, once stated his confusion about traditional marketing "Half the money I spend on advertising is wasted; the trouble is, I do not know which half." For a long time, traditional marketing has been regarded more as a means of consumption than a means of profit for business. Today, hundreds of years later, with the rapid development of the Internet and database technology, precision marketing emerges as time requires and gradually gains the favor of enterprises. First raised by Philip Kotler, precision marketing means sending messages only to the consumers who are the most relevant, at the most appropriate time, and through the exact channels. Unlike traditional marketing, which broadcasts the same message to a mass audience, precision marketing crafts different information for various sectors of a targeted audience. In the modern era, precision marketing is meaningful to both consumers and enterprises [1, 2].

Being able to access a wealth of information through the Internet, but with the same amount of time every day, current consumers have a shorter attention span for a piece of information than ever before. The intention of precision marketing is to push out the product information highly relevant to consumers' interests in an easily accessible way instead of waiting for them to find it among tons of irrelevant advertisements. Thus, precision marketing means to help consumers be more efficient in making every shopping decisions in their lives [3].

Consumers can do more things online, at the expense of traces of their personal information. To the extent permitted by law, consumer's behavior online will be collected in the database of enterprises. With big data, enterprises are empowered to construct accurate consumer profiles, define their most valuable customers, and predict their purchase behaviors. Though it requires an upfront expense on building a customer database, precision marketing is an investment for brands to achieve the most profitable result [4, 5]. Precision marketing can help companies to reduce costs and improve efficiency by reaching the most valuable customers more accurately. Compared to traditional marketing, precision marketing is more measurable; thus, enabling more optimization for future expenses. Precision marketing helps enterprises achieve long-term sustainable development by exploring new consumers, maintaining relationships with valuable consumers, and developing a loyal customer base for brands.

Because of its high reliance on big data, precision marketing is widely applied in business based on the Internet like e-commerce, streaming media, and social media. The e-commerce platforms like Amazon and Taobao employed precision marketing by showing consumers products that are similar to or related to the products they viewed or liked. In the retail industry, researchers emphasized the importance of providing personalized products and realizing an accurate marketing position in precision marketing. A study on WeChat, the Chinese social media app, proves that personalized precision marketing drives users' intention to click on ads [6–8].

Defined as crafting personalized messages for the most relevant customers, segmenting consumers is one of the most crucial parts of precision marketing. Traditionally, some most employed segmenting criteria include consumers' demographic (age, occupation, gender, etc.), psychographic, and behavioral characteristics. With the help of big data, various algorithms are employed in customer segmentation. Some of the methods include cluster analysis and backpropagation neural network [9].

2. Backpropagation Neural Network

The backpropagation neural network is a widely accepted and well-functioning neural network learning algorithm. The algorithm consists of two steps: the first step is the positive propagation process of the input sample. In the training stage, the network parameters are initialized randomly, including the connection weights and thresholds. The output of the input sample can be obtained after training of each layer of the network. Then, the output value is compared with the expected value. If the set error is not reached, the second step is carried out. The error signal is transmitted backward in the original path, and the weight value and the threshold value are continuously adjusted during the propagation process until the neural network error function or the number of iterations reaches the preset value to stop. The BP neural network generally consists of three layers as shown in Figure 1 [10–12].

The state of the neurons in each of these layers only affects the neurons in the next layer. According to the prediction error, the weights and thresholds of the network are continuously adjusted using backpropagation training. The output approximates the expected result until the predicted result meets the expected result. We set the input mode of the network to be $x = (x_1, x_2, \ldots x_n)^T$. The hidden layer has H units; the output of the hidden layer is $y = (y_1, y_2, \ldots y_h)^T$; the output layer has M units; their output is $z = (z_1, z_2, \ldots z_m)^T$, and the target output is $t = (t_1, t_2, \ldots, t_m)^T$. We set the transfer function of the hidden layer to the output layer to be f, and the transfer function to the output layer to be g, and we let the transfer function of the output layer be g. The output of the J th neuron in the hidden layer is as follows:

$$y_j = f\left(\sum_{i=1}^{n} w_{ij} x_i - \theta\right) = f\left(\sum_{i=0}^{n} w_{ij} x_i\right). \tag{1}$$
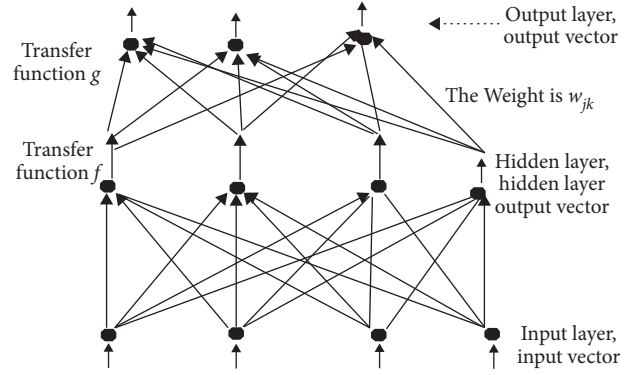

Figure 1: The BP neural network architecture.

Among them, $w_{0j} = -\theta$, $x_0 = 1 z_k = g(\sum_{j=0}^{h} w_{jk} y_j)$: output of the K th neuron in the output layer. Among them, $w_{0j} = -\theta$, $x_0 = 1$, output of the K th neuron in the output layer:

$$z_k = g\left(\sum_{j=0}^{h} w_{jk} y_j\right). \tag{2}$$

The error between the network output and the target output is $\varepsilon = (1/2) \sum_{k=1}^{m} (t_k - z_k)^2$. Obviously, it is a function of w_{ij} and w_{jk}.

The next step is to find a way to adjust the weight to make ε smaller. The BP neural network (backpropagation) adjusted sequence is carried out as follows:

(1) Adjust the weights from the hidden layer to the output layer. Set v_k to be the input $v_k = \sum_{j=0}^{h} w_{jk} y_j$ of neuron k in the output layer. Taking $g(x) = f(x) = (1/(1 + e^{-x}))$, the iterative formula for adjusting the weights from the hidden layer to the output layer is as follows:

$$g'(u_k) = \frac{e^{-v_k}}{(1 + e^{-v_k})^2} = \frac{1}{1 + e^{-v_k}}\left(1 - \frac{1}{1 + e^{-v_k}}\right) = z_k(1 - z_k). \tag{3}$$

(2) Weight adjustment from the input layer to the hidden layer.

$u_j = \sum_{i=0}^{n} w_{ij} x_i$, where u_j is the input of the J th neuron in the hidden layer:

The J th neuron in the hidden layer is connected with the neurons in the output layer. In other words, $\partial \varepsilon / \partial y_j$ involves all the weights. The weight from the input layer to the hidden layer is adjusted by iterating the following formula:

$$\frac{\partial \varepsilon}{\partial y_j} = \sum_{k=0}^{m} \frac{\partial (t_k - z_k)^2}{\partial z_k} \frac{\partial z_k}{\partial u_k} \frac{\partial u_k}{\partial y_j}$$
$$= -\sum_{k=0}^{m} (t_k - z_k) f'(u_k) w_{jk}. \tag{4}$$

3. Prediction of the User Behavior Label Based on the BP Neural Network

User label prediction is an abstract user model based on user network behavior or generated data. It is a model expression for a class of users with similar labels. Tagging users enables the computer to efficiently process user data, providing basic data support for precision marketing and enabling personalized recommendations for marketing activities or products. In user tagging prediction, first of all, we need to extract features based on the user's historical access data and then cluster the extracted user features to build a user tagging model. Finally, this model is used to predict the label of the user's personalized data and realize the accurate classification of users.

3.1. Expected Cross Entropy Feature Extraction Method for a User Text Vector. Users' history access data are usually presented in the form of text. Overcoming the challenge of how to extract useful features from massive text information is the key to user tagging. Feature extraction is the key technology of text classification. The quality of the feature subset extracted by feature will directly affect the effect of text classification. We select the feature which can effectively represent the whole information as the input source of the text classification model from the feature space. The key of the feature extraction is to find the optimal feature subset in the solution space that contains all the feature subsets and select the most representative feature combination under the premise of minimum time cost. The process of feature extraction can be summarized as follows: the original data set is preprocessed to get an initial feature set T, the feature set T is weighted, and then the feature set $T1$ is ranked according to the descending order of the weight values. According to the corresponding evaluation function, an optimal feature subset $T2$, which can best represent the information in the text category, is selected. The commonly used feature extraction methods include the Wrapper method [7], feature cluster selection algorithm based on distance measure [10], feature selection algorithm based on information measure [11], and so on. In this paper, first, a vector model of text space is established, in which each text is represented as a vector in the vector space, and each different feature item (entry) corresponds to a dimension in the vector space. The value of each dimension is the weight of the corresponding feature in the text. The vector space model represents text as a feature vector:

$$V(d) = ((t_1, w_1), (t_2, w_2), \ldots, (t_n, w_n)), \quad (5)$$

where $t_i \, (i = 1, 2, \ldots, n)$ is the featured item of document d, w_i is the weight of t_i, usually taken as a function of word frequency. Generally, the word is chosen as the feature of the document vector. The original vector representation is in the form of 0, 1. If the word appears in the text, then the dimension of the text vector is 1, otherwise 0. This approach fails to capture the word's effect in the text, thus gradually 0 and 1 are replaced by more accurate word frequencies. There are absolute and relative word frequencies. The absolute
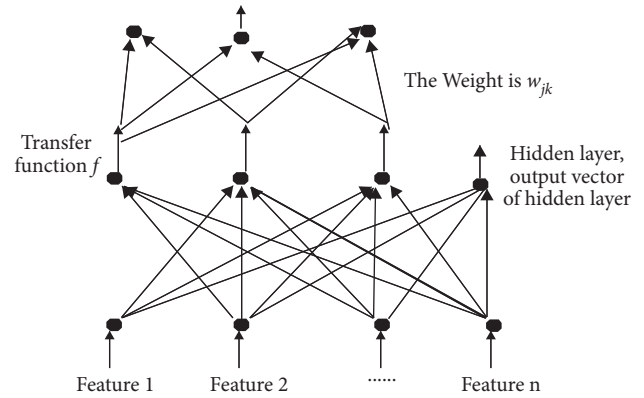

FIGURE 2: Two-layer prediction model of user tags.

word frequencies use the frequency of words appearing in the text to represent the text. The relative word frequency is converted into the normalized word frequency by TF-IDF.

After getting the text vector space, the expected cross-entropy is used to extract and analyze the text features. The cross-entropy is defined as follows:

$$H(X, q) = H(X) + D(p\|q) = -\sum_x p(x)\log(q). \quad (6)$$

It reflects the distance between the probability distribution of the text subject class and the probability distribution of the text subject class under the condition that a certain term appears. It differs only from information gain in that it does not take into account the fact that the word does not occur and only calculates the feature that appears in the text. If $p(C_i|W)$ is large and $p(CI)$ is small, then $p(CI)$ has a great influence on classification.

3.2. Prediction Model of User Label Based on the BP Neural Network. In order to achieve the accurate prediction of user tags, the classification model of feature extraction and the BP neural network prediction model are combined to form a two-layer prediction model of user tags as shown in Figure 2.

The BP neural network is used to train the user's text features and obtain the semantic information contained in the user's label. To ensure the integrity of information in feature construction and choose the appropriate parameters for the BP neural network, the experiment selects the nonlinear function as its activation function, enabling the neural network with the ability of nonlinear modeling and improving the versatility of its structure. If the activation function is a linear function, the multilayer neural network is equivalent to a simple multilayer perceptron with hidden layers and only has linear expression. The activation function generally constrains the input value within a certain range, which makes the gradient-based optimization method more stable. The common activation functions include sigmoid and Tanh, both of which are continuously differentiable. In the experiment, Tanh was chosen as the implicit layer activation function of the BP neural network and its convergence rate was faster than the sigmoid

TABLE 1: Description of dataset.

Name	Definition
User ID	Integer type, serialized user ID
Product ID	Integer type, serialized user ID
Product category ID	Integer type, the ID of the serialized item category
User behavior type	String, enumerated type, including ('pv', 'buy', 'cart', 'fav')
Timestamp	The timestamp of the event

TABLE 2: Types of user behaviors.

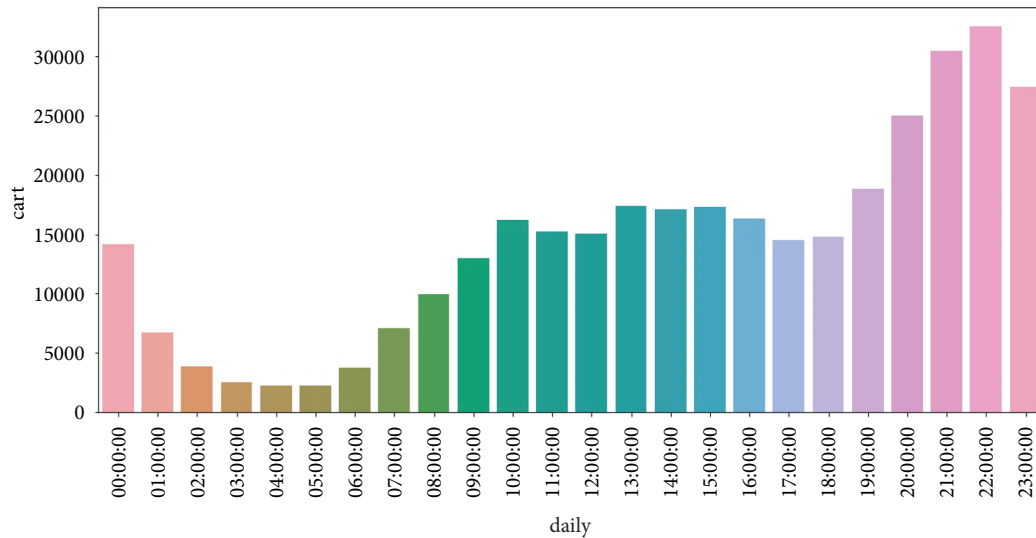| Behaviors	Definition
Pv	Page view of product details page, equivalent to click
Buy	Purchase products
Cart	Adding products into shopping cart
Fav	Adding products into collection/wish list

FIGURE 3: Adding products into the shopping cart.

function. We also choose SoftMax as the output layer activation function of the BP neural network, and the corresponding loss function is categorical cross entropy, for which the implicit layer transfer function f is Tanh.

4. Simulation

In order to test and verify the method in this paper, the User Behavior Data from Taobao for Recommendation dataset is used for BP network training and model building. This dataset contains all the actions (clicks, buys, adds, and likes) of approximately one million random users who engaged in behavior between November 25, 2017, and December 3, 2017. The dataset is organized in a form similar to MovieLens-20M, where each row of the dataset represents a user action, consisting of a user ID, a product ID, a product category ID, a behavior type, and a timestamp separated by commas. A detailed description of each column in the dataset is given in Tables 1 and 2.

The user behavior is analyzed and predicted according to different access times as shown in Figures 3–6.

As can be seen from Figures 3–6, the number of clicks, favorites, and add-ins to a shopping cart increased significantly during the night, but the increase was not as significant as other users' behavior during the night of purchase.

A cursory analysis of the purchase path without considering the sequence between collect and cart shows a user clicks on an item and then possibly collects it or adds it to the cart as a category. It is obvious that the buying rate rises from 8 a.m., peaks at noon, and then drops until 5 or 6 pm (seen in Figures 7-8). During working hours, the buying rate rises due to corporates demand. At this period of time the users perhaps often choose the thing they need. Thus, enterprises should pay attention to how to quickly satisfy users' needs during this time period. In the evening, the buying rate drops while the collection rate and the conversion rate into a shopping cart rises. While most users are at rest, it is important to emphasize what products might be added to users' carts, when they are not only focusing on necessities. It is
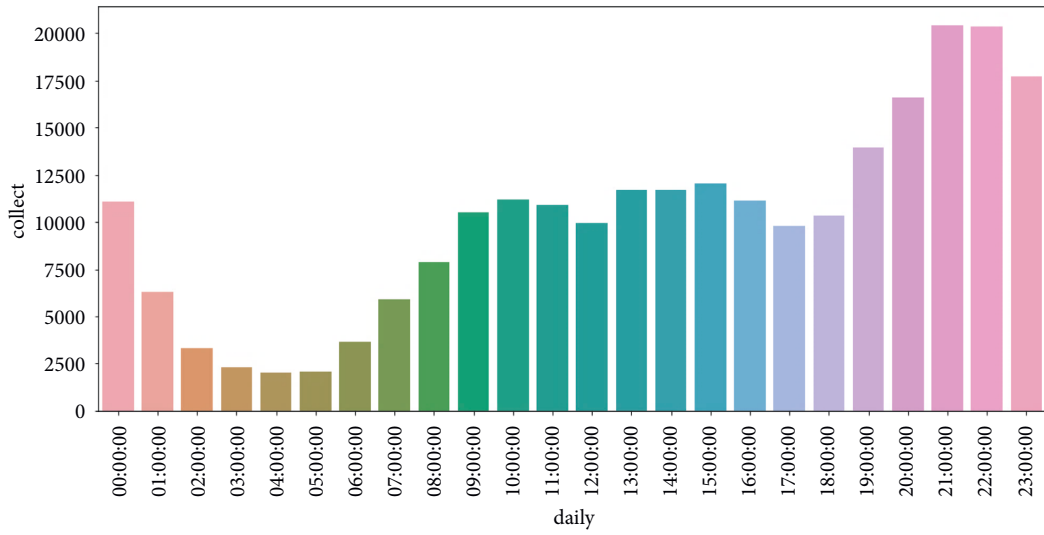
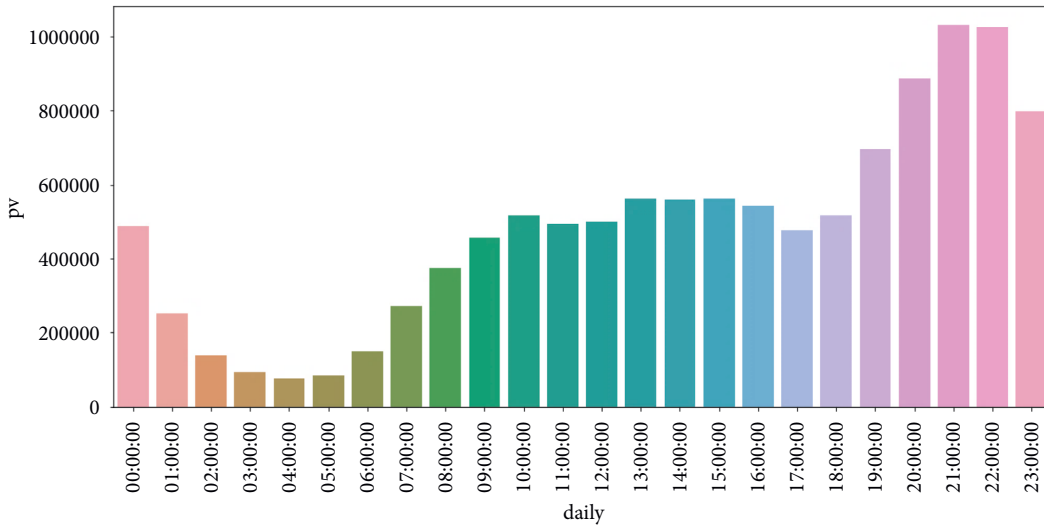FIGURE 4: Adding products into collection/wish list.

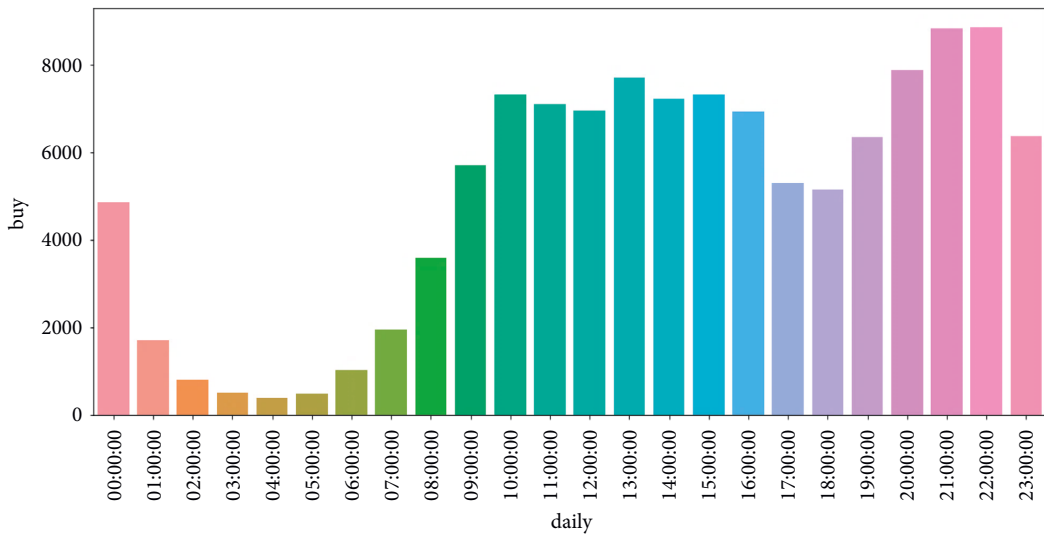FIGURE 5: Page view of product details page, equivalent to click.
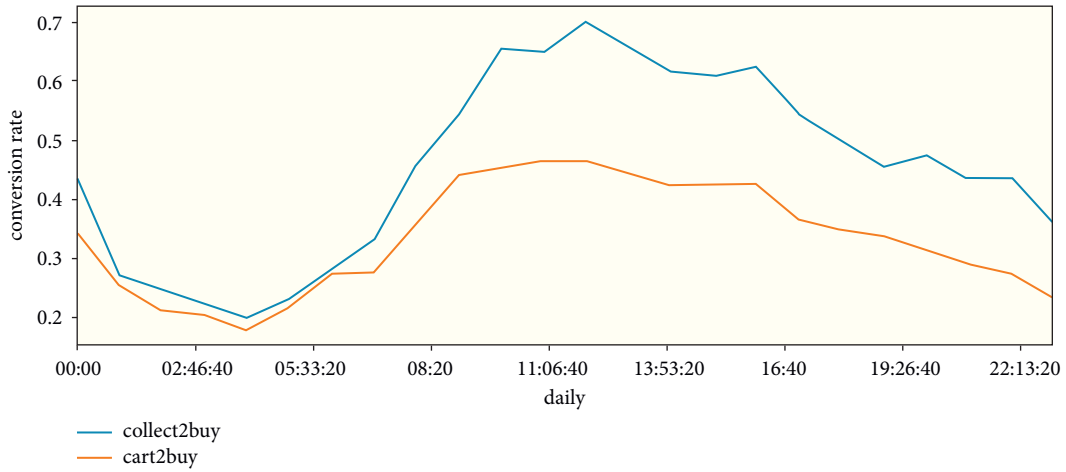

FIGURE 6: Purchase products.

FIGURE 7: Comparison of the conversion rate for collect2buy and cart2buy.
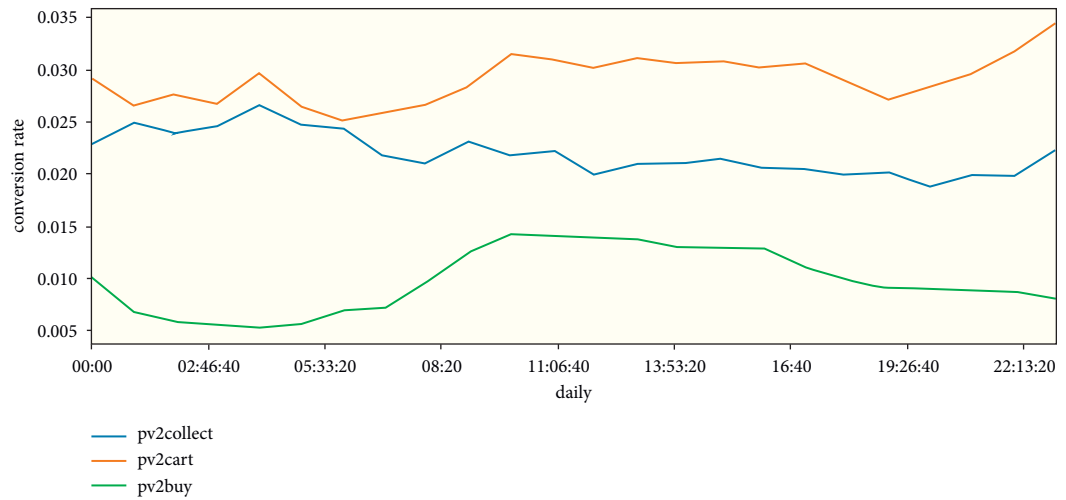

FIGURE 8: Comparison of the conversion rate for pv2collect, pv2cart, and pv2buy.
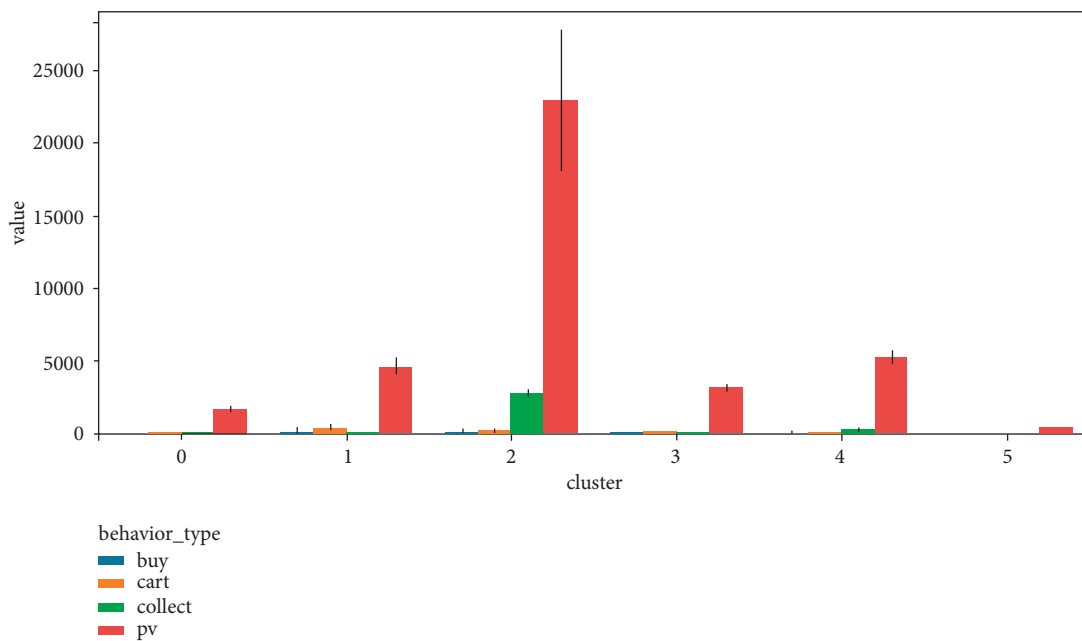

FIGURE 9: The value of users with different behaviors.

also necessary to study the buying habits of users during this period according to different commodity choices in order to find out the rules, so as to guide the attributes of commodities.

It can be seen in Figure 9 that the highest proportion of users in the first and last categories is 1,282 and 3,163, respectively, both of which belong to the groups with less user behavior. Also, the average purchase in the last category is much lower than the average. It lowers the overall average, with the purchase of the first group of users only more than the last group of users. From this, we can see that most users belong to this cluster of the user, which can be analogized with the 8 in the law of 2/8. There are 344 and 155 users in Category 4 and Category 5, respectively. The fourth type of user adds many products to the shopping cart but only collects a few, the fifth kind of user clicks and collects much but adds a few to their shopping cart. The last two groups with the lowest number of users were Category 2 and Category 3, with 56 and 2 users, respectively. The second group of users has more behaviors, and more importantly, their buying behavior is significantly higher than other types of users. In other words, this group of users are high-value ones, which are worthy to establish a strong relationship with, to maintain activity and stem the flow. The third category is more extreme, with a particularly high click-through rate but no matching purchase. We used a random sample of 5,000 users, and as the amount of data increased, the characteristics of each group became more pronounced. Such user groups can provide more information for business analysis, operation, and management. It can better support operators' refined management and scientific decision-making and can quickly catch the target users from the user behavior.

5. Conclusion

In summary, the study introduced the backpropagation neural network as a method for tagging users and achieving precision marketing. With the dataset from Taobao, the study employed the backpropagation neural network to process and analyze user behavior data, successfully labeled the users with tags, and summarized purchase patterns of different user clusters. In today's business environment, it is important for enterprises to know who their consumers are and the behavior pattern of different consumers in order to satisfy their needs. Targeting the most valuable users and sending them the most relevant messages are the keys to the success of the business.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] L. A. Yun, A. Yongyao, A. Chaawei et al., "Improving search ranking of geospatial data based on deep learning using user behavior data - ScienceDirect," *Computers & Geosciences*, vol. 142, 2020.

[2] L. Hu, Y. Xing, Y. Gong, K. Zhao, and F. Wang, "Nonnegative matrix tri-factorization with user similarity for clustering in point-of-interest," *Neurocomputing*, vol. 363, pp. 58–65, 2019.

[3] S. Konstantinidis, N. Moreira, R. Reis, and J. Young, "Regular expressions and transducers over alphabet-invariant and user-defined labels," *International Journal of Foundations of Computer Science*, vol. 31, no. 8, pp. 983–1019, 2020.

[4] C. Guan and Y. Yang, "Research on the application of back propagation neural network in the social development," *Computer Era*, vol. 5, pp. 46–52, 2021.

[5] H. Liu, H. Liu, J. Ruan, B. Tian, and S. Hu, "Research on precision marketing segmentation model based on Mining"Persona," *Journal of Silk*, vol. 52, no. 12, pp. 37–42, 2015.

[6] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, "Customer segmentation using K-means clustering," in *Proceedings of the 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, Udyambag, India, October 2018.

[7] W. Li and X. Jiang, "Research on the strategy of electronic commerce precise marketing," *Research and Development*, vol. 2, pp. 46–49, 2016.

[8] C. Yu, Z. Zhang, C. Lin, and Y. J. Wu, "Can data-driven precision marketing promote user ad clicks? Evidence from advertising in WeChat moments," *Industrial Marketing Management*, vol. 90, pp. 481–492, 2020.

[9] G. Zhu and X. Gao, "Precision retail marketing strategy based on digital marketing model," *Science Journal of Business and Management*, vol. 7, no. 1, p. 33, 2019.

[10] Y. Sun, N. Wang, C. Yin, and J. X. Zhang, "Understanding the relationships between motivators and effort in crowdsourcing marketplaces: a nonlinear analysis," *International Journal of Information Management*, vol. 35, no. 3, pp. 267–276, 2015.

[11] R. W. Cotterill, W. P. Putsis, and R. Dhar, "Market share and price setting behavior for private labels and national brands," *Food Marketing Policy Center Research Reports*, vol. 17, no. 1, pp. 17–39, 2000.

[12] L. Ableitner, V. Tiefenbeck, A. Meeuw, A. Worner, E. Fleisch, and F. Wortmann, "User behavior in a real-world peer-to-peer electricity market," *Applied Energy*, vol. 270, Article ID 115061, 2020.