Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 15, No. 2, 2024

Prediction Model of E-commerce Users' Purchase Behavior Based on Deep Learning

Ziqi Liu

Jinan Thomas School, Jinan, China

Abstract: This study mainly uses Deep Learning (DL) technology to build a prediction model of e-commerce users' purchasing behavior, and evaluates its application effect on actual e-commerce data. Methodologically, Recurrent Neural Network (RNN) is used to capture the temporal dependence of user behavior, and the performance of the model is improved through detailed data preprocessing and feature engineering. The research results show that the DL model based on RNN has achieved remarkable advantages in predicting the purchase behavior of e-commerce users. Compared with other methods, RNN model can capture the temporal dependence of user behavior more accurately, thus improving the prediction accuracy. This advantage is especially obvious when dealing with complex and dynamic user behavior data. It provides new ideas and methods for accurate marketing and personalized service of e-commerce industry.

Keywords: E-commerce, User purchase behavior, Deep Learning, Recurrent Neural Network, Prediction model.

1. Introduction

With the rapid development of Internet technology, e-commerce industry has gradually become an important part of the global economy [1]. From the initial books and costumes to home appliances and fresh food, the types of goods on the e-commerce platform are constantly enriched, and the scale of users is also constantly expanding [2]. However, in the fierce market competition, how to accurately capture user needs and predict user behavior has become the key to the sustainable development of e-commerce platforms. The importance of user's purchase behavior prediction is self-evident [3]. For the e-commerce platform, users' purchase intention can be predicted in advance, which can not only optimize inventory management, improve logistics efficiency, but also provide users with a more personalized shopping experience [4]. For example, through forecasting, the platform can recommend products that users may be interested in in advance, thus improving the conversion rate and user satisfaction [5].

In recent years, DL technology has made remarkable achievements in many fields. Compared with traditional forecasting methods, DL can better capture complex patterns and nonlinear relationships in data [6]. In the prediction of users' purchase behavior, DL can automatically learn the key factors that affect the purchase decision by using a large number of user behavior data, so as to predict the future behavior more accurately [7]. This advantage makes DL have a broad application prospect in the behavior prediction of e-commerce users.

The purpose of this study is to build a prediction model of e-commerce users' purchasing behavior by using DL technology. To this end, the article will first collect and sort out a large number of e-commerce user behavior data, including browsing records, purchase records, search records and so on. Then, the appropriate DL model will be selected, and the model will be trained and optimized. Finally, it is compared with other forecasting methods to highlight the advantages of the proposed method.

2. Theoretical Basis of E-Commerce Users' Purchase Behavior

2.1. Overview of Purchase Behavior of E-Commerce Users

The purchase behavior of e-commerce users refers to the behavior that users finally make purchase decisions through browsing, searching, comparing and selecting on the e-commerce platform [8]. This behavior not only involves users' personal preferences, needs and economic conditions, but also is influenced by many factors such as platform recommendation, product evaluation and promotion activities. The characteristics of e-commerce users' purchase behavior mainly include the following points: The user's behavior trajectory can be completely recorded and analyzed; the purchase decision of users is often influenced by a variety of information; Users' purchasing behavior is time-sensitive and dynamic, that is, it changes with the passage of time and changes in the market.

2.2. Influencing Factors of E-Commerce Users' Purchase Behavior

There are many factors that affect the purchase behavior of e-commerce users, among which price, quality and evaluation are the most critical factors. Price is one of the important factors that affect users' purchasing decisions, and users often weigh the price against the cost performance of goods according to their own economic conditions and needs. Quality is another important factor for users to consider. High-quality goods can win the trust and loyalty of users. In addition, users' purchasing decisions will be influenced by other users' evaluations, and positive evaluations can enhance users' purchasing willingness and confidence.

In addition to the above factors, there are many other factors, such as promotional activities, brand awareness, logistics and distribution, which will also affect users' buying behavior. These factors interact and influence each other to form a complex network of users' buying behavior.

2.3. The predicted Value of E-Commerce Users' Purchase Behavior

Accurately predicting the purchase behavior of e-commerce users is of great significance for e-commerce platforms. By predicting the purchase intention and demand of users, the platform can carry out inventory management and logistics distribution planning in advance, thus improving operational efficiency and reducing costs. At the same time, the forecast results can provide strong support for product recommendation and personalized service of the platform, and improve user experience and satisfaction. Through the prediction and analysis of purchasing behavior, the platform can also better understand the market dynamics and changes in user needs, and provide valuable reference information for future strategic planning and product development.

3. DL-Based Forecasting Model Construction

Before building a prediction model based on DL, adequate data preparation and pretreatment are crucial steps. This process mainly includes the following links:

Data source: This study collects user behavior data from several large e-commerce platforms, including browsing records, purchase history, search queries, etc. These data provide us with a wealth of user behavior information, which is the basis of building a prediction model.

Data cleaning: There are often problems such as missing values, abnormal values or duplicate values in the original data. Removing these noise data by data cleaning can ensure the accuracy and consistency of the data. For example, for missing values, we choose filling methods according to the data distribution, such as mean filling, median filling or predictive filling using machine learning algorithm.

Feature extraction: feature extraction is a key step in the pretreatment process. This paper extracts meaningful features from the original data, such as user browsing time, purchase frequency, search keywords, etc. These features will be used as the input of the model. At the same time, feature scaling, such as normalization or standardization, is also carried out to ensure that all features are on the same scale and further improve the performance of the model. In this paper, Mapminmax function is used to normalize the data. The expression of this function is as follows:

$$y = (y_{\max} - y_{\min}) \times (x - x_{\min}) \div (x_{\max} - x_{\min}) + y_{\min} \quad (1)$$

Where $y_{\max}$ and $y_{\min}$ are set constants, which are 1 and -1 respectively.

Data division: Finally, the data set after cleaning and feature extraction is divided into training set, verification set and test set. The training set is used to train the model, the verification set is used to adjust the model parameters and superparameters, and the test set is used to evaluate the performance of the model.

Considering that users' purchasing behavior is sequential, that is, users' purchasing decisions are often influenced by previous behaviors, this paper chooses RNN as the prediction model of this paper. RNN can process the sequence data and capture the dependencies in the sequence, which is suitable for predicting the purchase behavior of users. RNN is a special neural network structure, which processes sequence data through cyclic units. In RNN, the output of each time step depends not only on the current input but also on the state of the previous time step. This mechanism enables RNN to remember historical information and make predictions accordingly. The implementation steps of the prediction model based on RNN are as follows:

Determine input and output: the input is the characteristics of user behavior sequence, and the output is the prediction result of user purchase behavior (such as purchase/non-purchase).

Design RNN structure: select appropriate RNN units, and determine the number of hidden layers and units. The setting of these parameters needs to be adjusted according to the complexity of the problem and the amount of data. The output threshold of each neuron in the hidden layer is:

$$\{\theta_j\}, (j = 1, 2, \cdots, q) \quad (2)$$

Where r the threshold of neurons in the output is layer; n is the number of iterations. In this paper, the structure of RNN is shown in Figure 1.

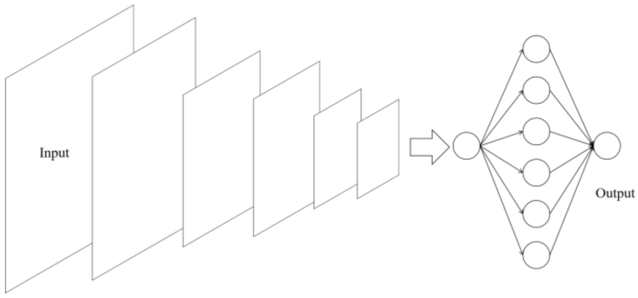

Figure 1. RNN structure diagram

Initialize weights and offsets: give initial values to the weights and offsets of the RNN model. These initial values can be assigned by random initialization or pre-training model.

Connection layer and activation function: add a full connection layer after the output layer of RNN, and use softmax function to convert the output into prediction probability.

Parameter setting: In the process of model construction, several parameters need to be set, including learning rate, batch size, iteration times, etc. The setting of these parameters has an important influence on the training speed and performance of the model. In this paper, these parameters are adjusted by experiments and the performance of the verification set to achieve the best training effect.

Model training is the core link in the process of DL model construction. In this paper, the loss function and optimization algorithm are defined to guide the training process of the model. (1) Definition of loss function: The loss function is used to measure the difference between the model prediction and the actual value. In this study, the binary cross-entropy loss function is chosen as the optimization objective, because it is suitable for the binary classification problem (purchase/non-purchase). (2) Selection of optimization algorithm: In order to minimize the loss function and update the weight and bias of the model, this paper chooses an appropriate optimization algorithm. Commonly used optimization algorithms include random gradient descent, Adam and so on. In this study, Adam optimization algorithm is chosen, because it combines the idea of momentum and RMSprop, and can usually converge faster and find a better solution.

In addition, in order to improve the prediction accuracy of the model, a series of tuning operations are carried out in this paper: the training speed and stability of the model are controlled by adjusting the learning rate and batch size; Use early stop technology to prevent over-fitting, that is, stop training in advance when the performance of verification set declines.

4. Model Application and Effect Evaluation

In order to verify the practicability and effectiveness of the model, this section applies the trained RNN model to the actual e-commerce data. Specifically, this section selects user behavior data in a period of time as input and predicts their future purchase behavior. By comparing with the actual purchase records, the forecasting ability of the model can be evaluated. The evaluation indexes include accuracy, recall and F1 score, and this method is compared with other prediction methods (logistic regression, decision tree and association rule mining), and the results are shown in Table 1.

Table 1. Performance comparison of prediction methods for purchase behavior of e-commerce users

Prediction technique	Accuracy (%)	Recall (%)	F1 score
RNN model	89.2	87.5	0.883
Logistic regression	76.3	72.1	0.742
Decision tree	79.4	75.0	0.771
Association rule mining	73.6	69.8	0.714

Remarks:

Accuracy: indicates the proportion of samples correctly predicted by the model to the total samples.

Recall: indicates the proportion correctly predicted by the model in real cases.

F1 score: it is the harmonic average of accuracy and recall, which is used to comprehensively evaluate the performance of the model.

Analysis of experimental results: As can be seen from the above table, the DL model based on RNN shows excellent performance in predicting the purchase behavior of e-commerce users. Its accuracy rate reached 89.2%, which means that nearly 90% of the purchase behavior was correctly predicted by the model. At the same time, its recall rate is as high as 87.5%, which shows that the model can capture most real buying behaviors. Combining the two methods, the F1 score of RNN model reached 0.883, which was significantly higher than other comparison methods.

Compared with other forecasting methods, such as logistic regression, decision tree and association rule mining, RNN model has obvious advantages in accuracy, recall and F1 score. This is mainly due to the fact that RNN model can effectively capture the temporal dependence of user behavior, thus showing higher prediction accuracy when dealing with complex and dynamically changing user behavior data.

5. Conclusions

In this study, the user's purchase behavior is deeply analyzed by constructing a prediction model based on DL. The experimental results show that DL model has significant advantages in dealing with complex user behavior data of e-commerce, and can effectively capture the potential patterns of user behavior. Specifically, the main findings and achievements of this study include: (1) DL model, especially RNN, performs well in predicting the purchase behavior of e-commerce users, and its prediction accuracy is obviously higher than that of traditional prediction methods. (2) Through detailed data preprocessing and feature engineering, the model can more accurately understand the dynamics and timing of user behavior, thus improving the accuracy of prediction. (3) It is also found that users' buying behavior is influenced by many factors, and DL model can automatically learn the importance of these factors without manual setting.

Although some achievements have been made in this study, there are still some shortcomings. In view of these shortcomings, the following aspects can be improved in the future: (1) expanding data sources and increasing the generalization ability of the model. (2) Try to combine other machine learning technologies, such as ensemble learning, to further improve the prediction accuracy. Looking forward to the future, DL will be more widely used in predicting the purchase behavior of e-commerce users. With the continuous progress of technology, more complex DL models will be developed to better understand and predict user behavior.

References

[1] Li Jin. Research on commodity demand forecast based on Alibaba's e-commerce data [J]. Journal of Zhejiang Wanli University, 2022, 35(2):85-91.

[2] Wang Tan, Zhou Qiyin, Mao Sha. Construction and application of the prediction model of shopping behavior of users of live broadcast e-commerce based on machine learning algorithm [J]. Journal of Hubei University: Natural Science Edition, 2023, 45(6):872-878.

[3] Yuan Weihua. Analysis of e-commerce sales data based on TensorFlow deep learning framework [J]. Computer Programming Skills and Maintenance, 2023(9):105-107.

[4] Cui Qing 'an, Wang Yaru. Research on the decision-making of social e-commerce users' consumption intention and purchase behavior in multi-dimensional context-an analysis of "Little Red Book" users as data collection objects [J]. Price Theory and Practice, 2020(12):95-98.

[5] Huang Chen. Analysis of e-commerce shopping user behavior based on big data [J]. Science and Technology Innovation, 2023(10):93-96.

[6] Zhou Yue, Zhou Jiu. User behavior prediction algorithm based on deep learning [J]. Digital Technology and Application, 2023, 41(10):154-156.

[7] Cui Teng. Emotional analysis of e-commerce users' comments based on deep learning [J]. Computer Knowledge and Technology, 2023(31):34-37.

[8] Liu Junyue, Ding Yi, Li Junfeng, et al. What determines the purchase intention of social e-commerce users? -Empirical explanation based on the theory of "cognition-emotion-willingness to act" [J]. Journal of Chongqing University of Technology: Social Sciences, 2022, 36(8):90-99.