SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(3), 562.
https://doi.org/10.62617/mcb562

Article

Research on quantitative measurement algorithm for e-commerce customer loyalty based on deep learning algorithm

Sian Chen

Zhejiang Institute of Communications, Hangzhou 310000, China; siachen42@zjvtit.edu.cn

CITATION

Chen S. Research on quantitative measurement algorithm for e-commerce customer loyalty based on deep learning algorithm. Molecular & Cellular Biomechanics. 2024; 21(3): 562.
https://doi.org/10.62617/mcb562

ARTICLE INFO

Received: 18 October 2024
Accepted: 28 October 2024
Available online: 19 November 2024

Abstract: Traditional algorithms cannot fully explore the potential patterns behind big data, lack personalized customer analysis, and cannot provide personalized services and suggestions for different types of customers. This article employs the Bi LSTM (Bidirectional Long Short-Term Memory) model to accurately capture the complex features and patterns of customer behavior, thereby improving the measurement accuracy of customer loyalty. Collect data on customer behavior, browsing history, and search behavior, and preprocess the collected data. Organize customer behavior data into a time series dataset in chronological order, and divide it into weekly windows to extract feature information from the data. Construct a bidirectional LSTM model while considering the forward and backward information of the sequence data, in order to more comprehensively capture the contextual relationships in the sequence data and quantify customer loyalty. The experimental results show that the average accuracy of Bi LSTM in predicting average customer loyalty is 97.1%. And it can effectively improve the prediction effect of repeat purchase rate. The application of Bi LSTM can accurately quantify customer loyalty in e-commerce, provide reference for enterprise decision-making, formulate corresponding marketing strategies and customer management plans, and improve customer loyalty and competitive advantages.

Keywords: e-commerce data; loyalty quantification; bidirectional long short-term memory; customer behavior

1. Introduction

The rise and development of e-commerce industry makes it more convenient for consumers to purchase products and services on the Internet [1,2]. However, with the intensification of competition, attracting and retaining customers has become increasingly challenging. Therefore, e-commerce companies are increasingly focusing on improving customer loyalty to ensure sustained profitability and business growth. Customer loyalty refers to the emotional connection and long-term purchase intention of customers towards a specific brand or enterprise, and is an important indicator for evaluating enterprise performance and long-term competitiveness [3,4]. The traditional methods for measuring customer loyalty in e-commerce are mainly based on statistical analysis and questionnaire surveys, but these methods have certain limitations. Statistical analysis requires a large amount of historical data, and questionnaire surveys may have issues with inaccurate information and sample bias. In recent years, with the rapid development of deep learning technology, customer loyalty quantification measurement methods based on deep learning algorithms have gradually received attention. Deep learning algorithms can learn complex patterns and patterns from a large amount of data, with stronger generalization ability and adaptability [5]. Therefore, applying deep learning algorithms to measuring customer loyalty in e-commerce has the potential to improve measurement and prediction accuracy. A customer behavior analysis model based on deep learning can more accurately capture customer purchasing preferences and behavioral habits, thereby more accurately evaluating customer loyalty levels.

Previous studies have shown that research on customer loyalty in e-commerce has mostly focused on using statistical methods or traditional machine learning algorithms for analysis. Customer loyalty research helps companies to gain a deeper understanding of customer needs, preferences, and behavioral habits, thus enabling better customer relationship management. Enterprises can enhance customer loyalty, establish long-term stable customer relationships, and achieve sustained business growth through personalized services and customized marketing strategies. In the research of customer loyalty quantification methods, traditional methods mostly rely on a single indicator of customer behavior and fail to capture the complex characteristics of customer behavior [6]. The Bi LSTM model can combine the forward and backward information of sequence data to effectively extract customer behavior patterns, enhance the accuracy of loyalty quantification, and provide data support for personalized marketing and customer relationship management. Al Ayed Sura used simple random sampling, and the data was collected through electronic questionnaires sent to study participants. These factors have a positive impact on establishing e-commerce customer loyalty—care, personality, choice, convenience, customization, and cultivation [7]. Khoa B aims to determine whether social media marketing can enhance the loyalty and online trust of electronic consumers. The results of this survey, which included 596 people, showed that social media marketing tools have a significant impact on consumer confidence and commitment to businesses through digital channels [8]. Akıl Siber aims to determine whether there is any relationship between the satisfaction and loyalty of e-commerce customers. The data of 1562 e-commerce customers living in Türkiye were collected through online survey. The results were analyzed using structural equation modeling [9]. However, these methods often overlook the temporal characteristics and complex correlations of customer behavior, and fail to fully explore the underlying patterns behind the data.

Deep learning can achieve more accurate and personalized recommendation systems by learning and analyzing a large amount of user behavior data. This system can provide customized product recommendations for users based on their historical purchase history, browsing behavior, preferences, and other factors, improving the shopping experience and user satisfaction, thereby promoting sales growth. Deep learning technology can analyze a large amount of user behavior data to provide personalized product recommendations for each user. By deeply understanding the purchasing history, preferences, and behavior patterns of users, personalized recommendation systems can improve user satisfaction and purchase conversion rates, thereby increasing the sales and profitability of e-commerce platforms [10,11]. Cao Yali conducts research from the perspective of establishing a financial warning model based on deep learning and constructing a financial risk warning mechanism for e-commerce companies, and analyzes and predicts the financial risks of listed companies. Through the construction of financial security early warning systems, crisis signals can be diagnosed as early as possible, and crisis signals can be prevented and resolved in a timely and effective manner [12]. Guo Lina proposed a model that combines convolutional neural networks and attention mechanisms to encode image features, and selected image features for products. A 5-layer convolutional neural network without fully connected layers was constructed to preliminarily extract image features, and then an attention mechanism strategy was designed [13]. Some researchers have used recurrent neural networks to capture the temporal characteristics of customer behavior, but there are still certain limitations when dealing with long-term dependency relationships. Bi LSTM has significant advantages in capturing customer behavior patterns. Compared with the one-way LSTM model, Bi LSTM can simultaneously utilize the forward and backward information of the customer behavior sequence to capture more comprehensive temporal characteristics. In the e-commerce scenario, by analyzing customers' browsing, purchasing and evaluation data through Bi LSTM, we can not only identify their current consumption preferences, but also explore potential needs and consumption tendencies. Case studies show that an e-commerce platform used the Bi LSTM model to predict customer repurchase behavior, and the accuracy increased by 15%, significantly better than traditional statistical models.

The aim of this study is to address the problems in quantifying customer loyalty in the field of e-commerce and propose a new method based on deep learning algorithms. Traditional loyalty measurement methods often only focus on a single behavioral indicator of customers, such as purchase frequency or consumption amount, while ignoring the complex characteristics and dynamic changes behind customer behavior. This method cannot comprehensively and accurately grasp the true situation of customer loyalty, which brings certain limitations to the marketing strategy formulation and customer relationship management of e-commerce enterprises. This article proposes a new method using a bidirectional long short-term memory model to address this issue. The Bi LSTM model, as a deep learning algorithm, has strong temporal modeling capabilities and can better capture the temporal characteristics and long-term dependencies of customer behavior. By combining forward and backward information, this model can comprehensively and dynamically analyze customer behavior, thereby improving the quantitative measurement accuracy of customer loyalty. Compared with traditional methods, algorithms based on the Bi LSTM model can more accurately predict customer future behavior and loyalty levels, providing more effective decision support and business optimization suggestions for enterprises.

2. Methods for quantifying loyalty measurement

2.1. Data collection and preprocessing

With the rapid development of e-commerce and increasingly fierce competition, attracting and retaining customers has become one of the key challenges for enterprises [14,15]. Customer loyalty is an important indicator for evaluating customer behavior and attitudes. Loyal customers tend to make long-term purchases and are more likely to recommend products or services to others, making them one of the most valuable assets of a business. By quantitatively measuring customer loyalty, enterprises can better understand their needs and preferences, provide more accurate products and services, and thus gain a competitive advantage.

Customer loyalty measurement is the core of customer relationship management.

By understanding the level of customer loyalty, enterprises can carry out targeted customer care activities, improve customer satisfaction, strengthen relationships with customers, promote customer repurchase and long-term cooperation [16,17].

By identifying customer groups with low loyalty, companies can optimize resource allocation, focus on and retain these customers, reduce customer churn rates and costs, and increase customer lifecycle value.

Loyalty measurement can help companies better understand the personalized needs and behavior patterns of customers, thereby implementing personalized marketing strategies [18,19]. By providing personalized services and discounts to customers with different levels of loyalty, customer satisfaction and loyalty can be enhanced.

Based on quantitative measurement of customer loyalty, enterprises can establish predictive models to predict future customer behavior and value. This helps companies take timely measures, adjust marketing strategies, retain customers with lower loyalty to the greatest extent possible, and promote the growth of high loyalty customers.

The quantitative measurement of customer loyalty in e-commerce has important background and significance in enhancing enterprise competitiveness, optimizing resource allocation, improving customer relationships, and implementing personalized marketing. It is one of the important basis for enterprises to formulate strategies and decisions. In order to effectively quantify loyalty, collect customer behavior data on e-commerce platforms, including purchase records, browsing history, search behavior, etc.

Collect data information from e-commerce websites, use Google Analytics for data collection, and collect user purchase records through the transaction system or order management system of the e-commerce platform. Record key information such as purchase time, product information, transaction amount, etc., and associate them with user identity information. Use data collection tools to track user browsing behavior on e-commerce platforms. Record the pages visited by users, browsing time, browsing depth, and other information to understand their interests and preferences. Use data collection tools to track user browsing behavior on e-commerce platforms. Record the pages visited by users, browsing time, browsing depth, and other information to understand their interests and preferences.

Table 1. Customer behavior data.

User number	Behavior type	Product number	Behavioral time
001	Buy	P001	2022-02-08 10:23:45
002	Browse	P002	2022-02-08 10:25:30
003	Search	P003	2022-02-08 10:26:15
004	Buy	P004	2022-02-08 10:27:00
005	Browse	P005	2022-02-08 10:28:20
006	Buy	P006	2022-02-08 10:30:10
007	Browse	P007	2022-02-08 10:31:45
008	Search	P008	2022-02-08 10:32:30
009	Browse	P009	2022-02-08 10:34:00
010	Buy	P010	2022-02-08 10:35:20

The collected customer behavior data is shown in **Table 1**.

By analyzing customer purchase records, browsing history, search behavior, etc., we can gain a deeper understanding of user behavior patterns, preferences, and habits, including the purchasing decision process, frequently purchased categories, browsing depth, etc.

The collected browsing history is shown in **Table 2**.

Table 2. Collected browsing history.

User number	Page path	Behavioral time	Browsing depth
1001	/products/electronics/laptops	2022-02-08 09:23	3
1002	/products/clothing/t-shirts	2022-02-08 10:45	2
1003	/products/home-appliances/vacuum-cleaners	2022-02-08 11:12	4
1004	/products/electronics/smartphones	2022-02-08 12:35	5
1005	/products/books/fiction	2022-02-08 13:48	3

In **Table 2**, each row represents a user's browsing behavior, including the user's ID, page path, visit time, and browsing depth (i.e., page depth or click count). These data can be used to analyze user browsing preferences, study user behavior patterns, and so on.

The collected search behaviors are shown in **Table 3**.

Table 3. Search behavior data table.

Search time	Search keywords	Search results click through status
2022-02-08 10:15:32	Mobile phone	Click on the first result
2022-02-08 11:30:45	Air purifier	Not clicked
2022-02-08 12:45:21	Television	Click on the third result
2022-02-08 14:20:10	Clothing	Click on the fifth result
2022-02-08 15:55:42	Jogging shoes	Click on the second result

In **Table 3**, each row represents a search behavior, including search time, search keywords, and click through status of search results. The search time records the specific time when the user searched, the search keywords record the search keywords entered by the user, and the click situation of the search results records whether the user clicked on the search results and the specific location of the click results.

Data cleaning and preprocessing are crucial steps in the data analysis process, ensuring the quality and consistency of data, and enabling subsequent analysis and modeling to be more accurate and reliable.

Check if there are any outliers in the search time, such as invalid date formats, time outside the scope of data collection, etc. Search time can be checked through data visualization or programming methods, and non-compliant time records can be considered as outliers and deleted or corrected.

Check for abnormal values such as null values, illegal characters, duplicate values, etc. in search keywords and search results clicks. Search keywords and click through status can be checked through data statistical analysis, and outliers can be handled.

In search behavior data, there may be missing search keywords or search result clicks. For the processing of missing values, fill in with the most frequent keywords. For search result clicks, missing values can be considered as not clicked and filled in. If there are other related fields in the dataset, missing values can also be filled in through the information of the relevant data, such as inferring missing search keywords or clicks based on the user's other search behaviors.

In the data preprocessing process, it is also necessary to standardize the data so that different features have the same scale and range, avoiding the impact of differences between features on model training. For search behavior data, it may be necessary to standardize the search time, such as dividing the date and time into different dimensions such as year, month, day, hour, minute, and second for standardization. For other features, such as the length of search keywords and the classification of search result clicks, standardization can also be applied to make the data have similar scales and ranges, which is beneficial for subsequent modeling and analysis.

The formula for data standardization is:

$$z = \frac{x - u}{\sigma} \tag{1}$$

In the process of data cleaning and preprocessing, it is also necessary to verify the consistency of the data to ensure that it conforms to the expected data patterns and structures. For example, verifying whether the click through status of search results meets predefined classification criteria, verifying whether the search time meets the expected time format and range, and verifying whether the data maintains consistency after filling in missing values.

2.2. Building a time series dataset

Organizing customer behavior data in chronological order into a time series dataset is to better understand and analyze the trend of customer behavior over time. When organizing time series datasets, an appropriate time window size can be set to partition the time series. The size of the time window determines the granularity of time series data, and dividing it by week can clearly observe long-term trends and periodic changes.

Organize customer behavior data into a time series dataset in chronological order. Firstly, it is necessary to sort the customer behavior data based on the temporal information, ensuring that the data is arranged in chronological order.

Then, based on the selected time window size, the data is divided into multiple time windows, and the customer behavior data within each time window is organized into a time series data point.

Each time series data point consists of a timestamp and corresponding customer behavior data. The timestamp represents the start or end time of the time window to which the data point belongs, while customer behavior data includes purchase records, browsing history, search behavior, etc.

You can choose different timestamp representations according to your needs, such as using the start time of the time window as the timestamp, or using the middle or end time of the time window as the timestamp.

When organizing time series datasets, there may be situations where there is no customer behavior data in some time windows, resulting in missing values. For the processing of missing values, you can choose to delete or fill in the missing values, depending on the sparsity of the data and the needs of analysis.

If there are fewer missing values that do not affect the overall analysis results, you can choose to delete the missing values; If there are many missing values or the distribution pattern of missing values has certain characteristics, filling in the missing values can be chosen, such as using the interpolation mean method for filling.

In the process of organizing time series datasets, it is necessary to verify the consistency of the dataset and ensure that the format and content of data points within each time window meet the expected requirements. Data can be visualized or subjected to descriptive statistical analysis to view the distribution and characteristics of the data.

The process of constructing a time series dataset is shown in **Figure 1**.

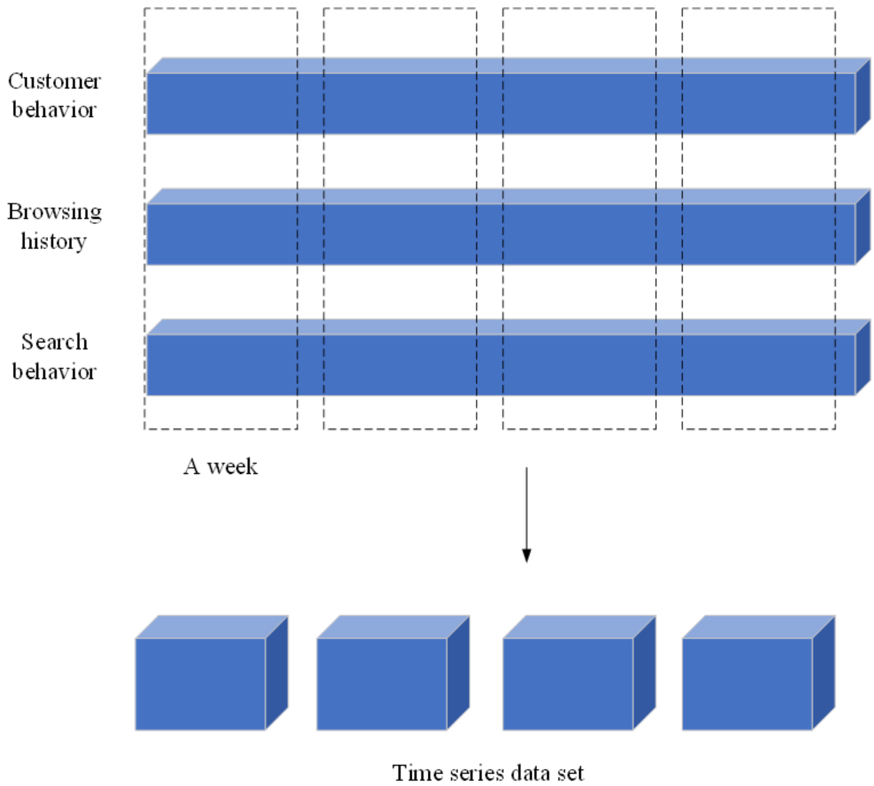

Figure 1. Process of constructing a time series dataset.

Organizing data according to one week can provide a clearer observation of customer behavior patterns during the week. Different behaviors may exhibit different trends and patterns at different times of the week, such as higher weekend shopping volume than weekdays, or higher nighttime browsing volume than daytime.

2.3. Feature engineering

Extracting features from customer behavior data is to better understand customer behavior patterns and characteristics, thereby providing a reliable data foundation for subsequent analysis and modeling.

Purchase frequency refers to the number of times customers make purchases within a certain period of time. The purchase frequency can be calculated by counting the number of purchases made by each customer within a week, a month, or a year. Calculate the purchase frequency of each customer based on timestamps and count the number of purchases made by each customer within each time window.

The purchase amount refers to the amount paid by customers when purchasing goods or services. The purchase amount feature can be calculated by counting the purchase amount of each customer over a certain period of time. Directly use the purchase amount information in the database to calculate the purchase amount for each customer.

The number of views refers to the number of times a customer views a product or page while visiting an e-commerce platform. The browsing frequency feature can be calculated by counting the number of times each customer views within a certain period of time.

Search keywords refer to the keywords that customers enter when searching on e-commerce platforms. Search keyword features can be extracted by counting the search keywords of each customer over a certain period of time.

In the process of extracting features, it is necessary to design and select appropriate features based on business needs and actual situations. The selection of features should consider factors such as correlation between features, contribution to business objectives, and interpretability of features. At the same time, it is necessary to preprocess and standardize the features to ensure that they have the same scale and range for subsequent data analysis and modeling.

Feature engineering refers to the transformation and extraction of raw data in machine learning and data analysis to create new features or adjust existing features, thereby improving the performance and effectiveness of the model. Xiu's feature engineering can reduce the complexity of the model and reduce the risk of overfitting. By selecting features with high correlation and low redundancy, the training time and computational resources of the model can be reduced, and the model can be made more generalizable.

Feature engineering can make models more interpretable and help understand the predictive results of the model. By processing and transforming features, they can be made more in line with actual business scenarios, thereby enhancing the interpretability of the model. Feature engineering can make data more visually appealing, aiding in data analysis and decision-making. By extracting and creating new features, data can be presented more intuitively in charts and visualization tools, helping users better understand data features and trends.

Feature engineering can mine potential information and hidden patterns in data, thereby enriching the expressive power of the data. By extracting and transforming features from data, it is possible to make it more representative and interpretable, discover patterns and trends in the data, and provide more reference basis for decision-making.

Feature engineering plays a crucial role in machine learning and data analysis. By designing and selecting features reasonably, model performance can be improved, model complexity can be reduced, model interpretability can be improved, data visualization can be enhanced, different models and algorithms can be adapted, and potential information can be mined, providing more powerful support for data analysis and decision-making.

2.4. Building a bidirectional LSTM model

Long short-term memory network is a commonly used recurrent neural network, mainly used for processing and predicting time series data. Compared to traditional recurrent neural networks, LSTM has a better ability to capture long-term dependencies in time series data, making it suitable for various sequence modeling tasks. The LSTM model consists of a series of LSTM units, each containing three gates (input gate, forget gate, and output gate), as well as a memory cell for storing and transmitting information.

However, the unidirectional LSTM model still has shortcomings. The STM model can only process sequence data unidirectionally along the timeline, so it cannot directly capture bidirectional contextual information in sequence data. This means that in some cases, LSTM may miss some important contextual information, leading to a decrease in the performance of the model.

Although LSTM models have certain long-term memory capabilities, there is still a problem of insufficient long-term dependency modeling ability when dealing with particularly long sequence data. In this case, LSTM may encounter problems such as vanishing or exploding gradients, making it difficult to train the model or slow convergence during training.

The LSTM model requires sequential and gradual updating of hidden states and memory cells when processing sequence data, which leads to low computational efficiency, especially when dealing with long sequence data. Due to the fact that the calculation of each time step depends on the result of the previous time step, parallelization cannot be performed, resulting in relatively slow training speed.

LSTM models typically require fixed length sequence data as input, making it relatively difficult to process variable length sequence data. In practical applications, if the length of the sequence data is inconsistent, it needs to be filled or truncated, which may affect the performance and generalization ability of the model.

There are multiple important hyperparameters in the LSTM model that need to be adjusted, including hidden layer size, number of memory cells, learning rate, etc. Adjusting these hyperparameters requires a certain amount of experience and experimentation, and the optimal parameter settings may vary for different datasets and tasks, thus requiring extensive experimentation and tuning work.

The bidirectional LSTM model can learn and capture information from both directions of the sequence simultaneously, thereby gaining a more comprehensive understanding of the bidirectional contextual information in sequence data [20,21]. This enables the model to better understand the context and correlations in sequence data, thereby improving its modeling ability for sequence data. The bidirectional LSTM model is shown in **Figure 2**.

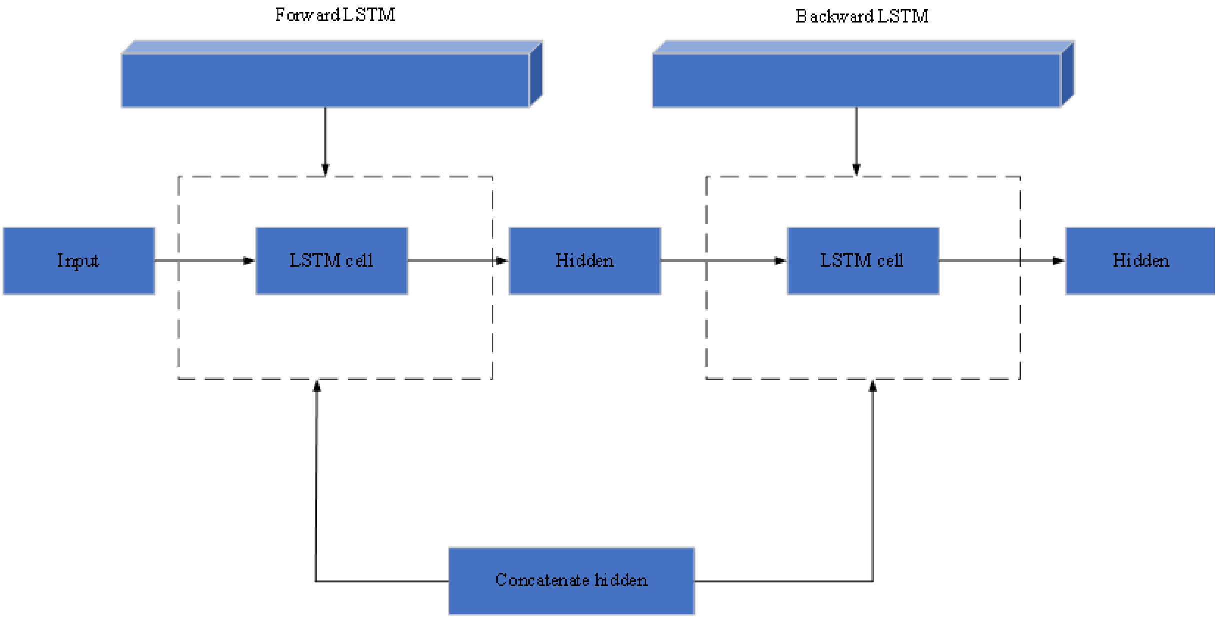

Figure 2. Bidirectional LSTM model.

In the bidirectional LSTM model, the input sequence data is processed simultaneously through LSTM units in two directions, namely forward LSTM and backward LSTM. These two LSTM units have the same structure, but handle the forward and backward information of the input sequence data separately.

The advantage of bidirectional long short-term memory networks is that they can simultaneously consider the forward and backward information of sequence data, thereby more comprehensively capturing contextual relationships in sequence data, improving the model's expressive power and predictive performance. The bidirectional LSTM model can effectively capture long-term dependencies in sequence data, including dependencies between distant positions in the sequence. This enables the model to have stronger modeling capabilities when processing various sequence data, such as sentence comprehension in natural language processing and speech recognition in speech recognition.

The bidirectional LSTM model can effectively handle variable length sequence data, where the length of each sample in the sequence data is inconsistent. This makes the model more flexible and applicable in handling sequence data in various practical applications. The bidirectional LSTM model can better capture contextual relationships in sequence data, thereby improving the model's generalization ability. When the training and testing sets have different contextual features, the bidirectional LSTM model can better adapt to different data distributions, thereby improving the model's generalization performance.

Using sigmoid and tanh functions as activation functions, sigmoid and tanh functions are used in gating units to control the threshold and amplitude of information flow. The formula for the sigmoid function is:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

The formula for the Tanh function is:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

In the Bi LSTM model, due to its bidirectional characteristics and a large number of parameters, model complexity may lead to overfitting. Through regularization techniques, L2 regularization effectively manages the complexity of the model. By adding penalty terms to limit the size of the weights, the degree of freedom of the model is reduced to prevent overfitting. The dropout technique can randomly discard some neurons during the training process, further enhancing the generalization ability of the model.

The attention mechanism dynamically assigns different weights to different parts of the input sequence, allowing the model to focus on the information most relevant to the current prediction. This approach increases the model's sensitivity to key features, thereby improving the accuracy and interpretability of the prediction, making the model's decision-making process more transparent and understandable.

The significance and role of quantifying customer loyalty lies in helping enterprises gain a deeper understanding of customer behavior, improve customer satisfaction, promote customer retention and growth, enhance brand loyalty, and optimize marketing strategies. By quantifying customer loyalty, companies can gain a deeper understanding of customer behavior habits, purchasing preferences, consumption frequency, and other information. This helps enterprises to more accurately grasp customer needs and preferences, providing reference for product development and service optimization.

By quantifying customer loyalty, companies can promptly identify and address customer dissatisfaction, improve product quality, service level, and shopping experience, thereby enhancing customer satisfaction and loyalty. Understanding customer loyalty levels can help businesses develop targeted customer retention strategies, such as offering discounts and customized services, thereby reducing customer churn and promoting customer retention and growth.

By quantifying customer loyalty, companies can better understand their perception and emotional attitudes towards the brand, thereby creating and maintaining a targeted brand image, and enhancing customer trust and loyalty towards the brand. By quantifying customer loyalty, enterprises can continuously improve their products and services, enhance brand competitiveness, and stand out in fierce market competition, gaining greater market share and revenue.

Customer loyalty quantification measures the degree of customer loyalty to a brand or enterprise through indicators such as customer behavior and experience. The repeat purchase rate and retention rate are commonly used indicators to measure customer loyalty.

The formula for quantifying customer loyalty is:

$$L = w_1 \times A + w_2 \times B \tag{4}$$

In Equation (4), A and B are the repeat purchase rate and retention rate, respectively, and w_1 and w_2 are the corresponding weights.

3. Quantitative measurement and evaluation of customer loyalty

3.1. Model training

The purpose of quantifying customer loyalty is to help enterprises gain a deeper understanding of their loyalty to their brand or products, in order to better meet customer needs, improve customer satisfaction, promote customer retention and growth, and ultimately achieve sustained and stable business growth. By quantifying customer loyalty, companies can evaluate and measure customer recognition and trust in the brand from multiple dimensions such as customer behavior and experience, and then develop corresponding strategies and measures to improve customer loyalty. At the same time, quantifying customer loyalty can also help enterprises identify and explore high-quality customers, improve their lifecycle value, and thus achieve higher profits and market competitiveness.

This article uses a bidirectional LSTM model for quantitative analysis of customer loyalty. In order to effectively measure and evaluate customer loyalty, cross validation is used for dataset partitioning. And improve the quantitative effect of customer loyalty by training the bidirectional LSTM model.

Divide the preprocessed and feature engineering processed dataset into training sets. The training set should include an input sequence and corresponding labels, where the input sequence can be time series data or other sequence data, and the label can be the value of the next time step of the sequence or a classification label. After the model construction is completed, it is necessary to select the appropriate loss function and optimizer for model training. Using mean square error as the loss function and using Adam optimizer.

Train the bidirectional LSTM model using the prepared training set. During the training process, the model will continuously adjust its parameters through backpropagation algorithm to minimize the loss function. During the training process, some hyperparameters can be set, with a batch size of 32 and 200 training rounds, to control the training process of the model.

The grid search technique is used in the hyperparameter selection process. First, define the hyperparameters to be tuned and their value ranges, such as learning rate, batch size, and number of network layers. Then, create a parameter grid containing all possible combinations. Next, cross-validate each combination to evaluate its performance on the validation set. Finally, compare the results of all combinations and select the hyperparameter combination with the best performance. The advantage of this method is that it is systematic and comprehensive, and can effectively find the best parameter configuration, thereby improving the accuracy and stability of the model.

3.2. Evaluation indicators

By predicting customer loyalty, enterprises can better understand their needs and preferences, adjust products and services in a timely manner, improve customer satisfaction and experience, and thus promote the improvement of customer satisfaction. Understanding customer loyalty levels can help businesses develop more precise marketing strategies, such as launching different marketing activities based on customers with different levels of loyalty, improving marketing efficiency and return on investment.

Customer loyalty prediction is of great significance for the development and competitiveness of enterprises. It can help enterprises better understand customer needs, reduce customer churn, optimize marketing strategies, improve brand loyalty, and achieve long-term stable business growth.

In order to effectively evaluate the effectiveness of customer loyalty prediction, the evaluation indicators set include accuracy, precision, and recall. The accuracy formula for predicting customer loyalty is expressed as:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

The formula for precision is:

$$P = \frac{TP}{TP + FP} \tag{6}$$

The formula for recall rate is:

$$R = \frac{TP}{TP + FN} \tag{7}$$

In order to analyze the effectiveness of the bidirectional LSTM model in predicting customer loyalty more effectively, the bidirectional LSTM is compared with LSTM, SVM (Support Vector Machine), and RF (Random Forest). And analyze the effectiveness of predicting repeat purchase rates under different models.

4. Results

4.1. Customer loyalty prediction accuracy

The role of customer loyalty prediction is to help enterprises better understand and predict customer behavior, identify potential loyal customers and lost customers in advance. The accuracy results of different models for customer loyalty prediction are shown in **Figure 3**.

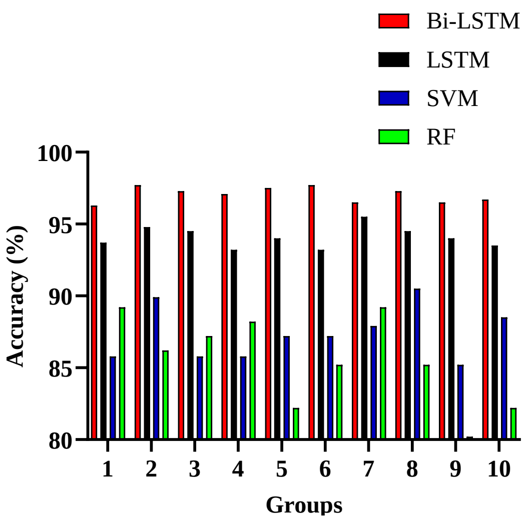

Figure 3. Accuracy of customer loyalty prediction.

The accuracy of customer loyalty prediction using the bidirectional LSTM model is higher than that of LSTM, SVM, and RF. The average prediction accuracy of Bi LSTM, LSTM, SVM, and RF are 97.1%, 94.1%, 87.4%, and 85.5%, respectively. Bidirectional LSTM can capture the temporal information of input sequence data and consider both forward and backward information, thereby gaining a more comprehensive understanding of customer behavior patterns and trends.

The accuracy of customer loyalty prediction lies in the model's ability to deeply explore and analyze the temporal characteristics of customer behavior. The bidirectional LSTM model outperforms other models (LSTM, SVM, and RF), mainly due to its unique network structure, which can process both forward and backward information of input data. This bidirectional processing method makes the model more comprehensive in understanding customer behavior patterns and can capture potential trends and changes. When analyzing customer purchase history, the bidirectional LSTM can not only consider current and past behaviors, but also reversely examine future possibilities, thereby providing more accurate predictions. In addition, the bidirectional LSTM performs well in processing long sequence data and can effectively alleviate the gradient vanishing problem that may occur in traditional LSTM when processing long dependencies. In contrast, models such as SVM and RF lack flexibility in processing nonlinear and time series data, resulting in relatively low prediction accuracy. Therefore, the bidirectional LSTM not only improves the accuracy of customer loyalty prediction, but also provides companies with deeper customer insights and helps to formulate more targeted marketing strategies.

4.2. Customer loyalty prediction accuracy

The accuracy results of customer loyalty prediction using different models are shown in **Figure 4**.

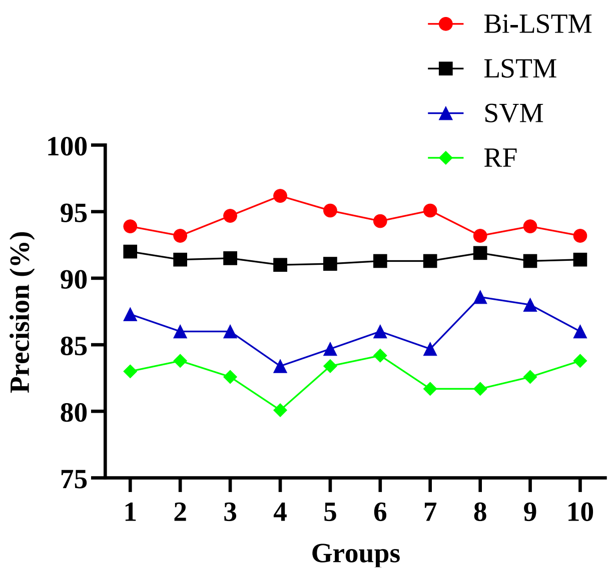

Figure 4. Accuracy of customer loyalty prediction.

Customer behavior data often has sequence characteristics, such as purchase history, browsing history, etc. Traditional classifiers such as SVM and RF cannot directly process sequence data, while LSTM and Bi LSTM can effectively process this type of sequence data. The bidirectional LSTM model has more parameters and stronger expressive power, which can better adapt to complex customer behavior patterns and data features, thereby improving the prediction accuracy of the model.

4.3. Customer loyalty prediction recall rate

The recall rate results of customer loyalty prediction using different models are shown in **Figure 5**.

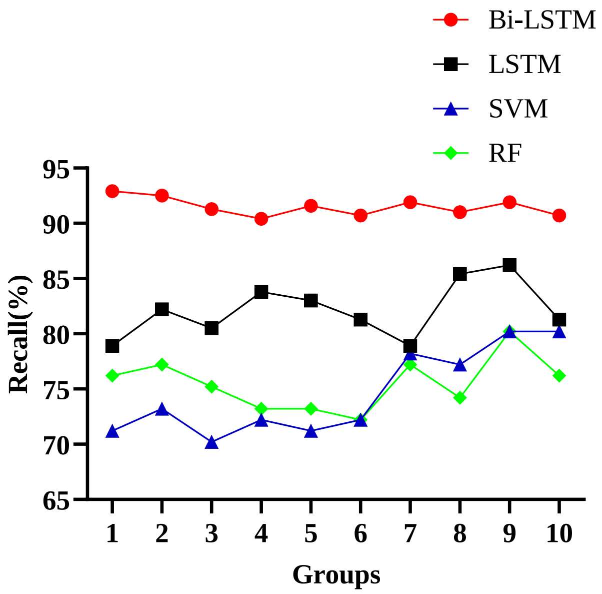

Figure 5. Customer loyalty prediction recall rate.

LSTM and Bi LSTM have adaptive learning ability, which can dynamically adjust model parameters according to changes in data, adapt to different data distributions and features, and thus improve the generalization ability of the model. The bidirectional LSTM model can achieve end-to-end learning, directly learning feature representations from raw data, avoiding the process of manual feature extraction, and reducing the complexity and subjectivity of feature engineering. The bidirectional LSTM model has a higher recall rate in customer loyalty prediction compared to traditional models such as LSTM, SVM, and RF, mainly because it can effectively process temporal information and sequence data, and has stronger expressive and adaptive learning abilities.

4.4. Prediction accuracy of repeat purchase rate

The prediction of repeat purchase rate can help enterprises identify customers with high repeat purchase potential in a timely manner, and adopt targeted marketing strategies. The accuracy of repeat purchase rate prediction using different models is shown in **Table 4**.

Bi LSTM can effectively capture the temporal information of input sequence data, including purchase history, browsing history, etc. However, traditional models such as SVM and RF cannot directly process sequence data, while LSTM can only consider forward information. Considering that temporal information helps the model better understand customer behavior patterns and trends, Bi LSTM has an advantage in predicting repeat purchase rates. By predicting the repeat purchase rate of customers, enterprises can promptly identify potential customers that may be lost and take measures to prevent economic loss. This can reduce the cost of customer acquisition, improve customer retention, and thus reduce marketing costs and resource waste.

Table 4. Prediction accuracy of repeat purchase rate.

Group	Bi-LSTM	LSTM	SVM	RF
1	93.4	83.2	75.4	67.5
2	94.5	87.6	73.4	73.4
3	95.8	88.2	68.5	74.4
4	94.2	87.6	70.5	70.5
5	95.8	84.9	67.5	72.4
6	95.6	85.4	74.4	70.5
7	96.1	88.7	74.4	68.5
8	93.9	84.9	72.4	74.4
9	95.8	87.6	76.4	68.5
10	94.5	85.4	72.4	73.4

In order to enhance the generalization ability of the Bi LSTM model, especially its applicability to different e-commerce platforms and product types, by introducing multi-task learning and transfer learning techniques, the model can effectively transfer the feature knowledge learned on one platform to another platform, reducing the dependence on a large amount of labeled data. For different types of products, by analyzing the unique behavior patterns of various products, the input features are adjusted to make the model more targeted. During the training process, data enhancement technology is used to further improve the model's adaptability to new data. This method not only improves the performance of the model on different e-commerce platforms, but also enhances its recognition ability on various product types, ensuring the stable performance of the model in a dynamic market environment, thereby achieving more accurate customer behavior prediction.

5. Conclusion

Understanding the changing trends and influencing factors of customer loyalty can help businesses more accurately formulate marketing strategies and promotional activities. This article uses a bidirectional LSTM model to predict customer loyalty, collecting customer behavior data from e-commerce platforms, including purchase records, browsing history, search behavior, etc. The bidirectional LSTM model consists of two LSTM layers, forward and backward, each of which may contain multiple LSTM units. By adjusting the parameters and structure of the model, it is ensured that the model can fully explore the information in customer behavior data and improve prediction accuracy. The experimental results show that the bidirectional LSTM model can better improve the accuracy of customer loyalty prediction compared to LSTM, SVM, and RF. By quantitatively measuring customer loyalty, enterprises can better understand their attitudes and loyalty towards their brand or product, and thus develop more accurate customer relationship management strategies.

However, the data collected in this article may not be comprehensive enough, and collecting more comprehensive e-commerce data for analysis in the future will be the direction of future research.

Funding: This research is supported by China Vocational Education Association of Zhejiang Province grant number [No. ZJCV2024C01].

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

Reference

1. Prajapati, Dhirendra, Felix T. S. Chan, Yash Daultani & Saurabh Pratap. "Sustainable vehicle routing of agro-food grains in the e-commerce industry." International Journal of Production Research 60.24 (2022): 7319-7344.
2. Febransyah, Ade, and Joklan Imelda Camelia Goni. "Measuring the supply chain competitiveness of e-commerce industry in Indonesia." Competitiveness Review: An International Business Journal 32.2 (2022): 250-275.
3. Pan Mengshi, Bai Huanhuan, and Li Pei "The impact of after-sales service quality of clothing online stores on customer loyalty." Journal of Beijing Institute of Fashion 42.1 (2022): 49-54
4. Jiang Yuzhen, Lin Junxing, and Li Hongchao "A Study on the Relationship between Store Image, Customer Satisfaction, and Customer Loyalty - Taking Starbucks as an Example." Policy and Human Resources Management 13.1 (2022): 1-38
5. Yanling Jiang "Analysis of the Impact and Countermeasures of E-commerce on Business Administration." International Academic Forum on Finance and Management 2.7 (2024): 4-6
6. Tao Qu "Research on International Trade and E-commerce Strategies in the Fine Chemicals Industry." Finance and Management International Academic Forum 2.7 (2024): 31-33
7. Al-Ayed, Sura. "The impact of e-commerce drivers on e-customer loyalty: Evidence from KSA." International Journal of Data and Network Science 6.1 (2022): 73-80.
8. Khoa, B., and T. Huynh. "The influence of social media marketing activities on customer loyalty: A study of e-commerce industry." International Journal of Data and Network Science 7.1 (2023): 175-184.
9. Akıl, Siber, and Mustafa Cahit Ungan. "E-commerce logistics service quality: customer satisfaction and loyalty." Journal of Electronic Commerce in Organizations (JECO) 20.1 (2022): 1-19.
10. Xu Ruiqing, Zhang Zhiwang, and Sun Hongliang "A New Method for E-commerce Water Army Detection: Multi relation Graph Neural Network with Adaptive Neighborhood Precision Sampling." Library and Information Knowledge 39.6 (2023): 35-44
11. Mao Yuanhong, Sun Chenchen, Xu Luyu, Liu Xi, Chai Bo, He Pengchao, et al. "A Review of Time Series Prediction Methods Based on Deep Learning." Microelectronics and Computer 40.4 (2023): 8-17
12. Cao, Yali, Yue Shao, and Hongxia Zhang. "Study on early warning of E-commerce enterprise financial risk based on deep learning algorithm." Electronic Commerce Research 22.1 (2022): 21-36.
13. Guo, Lina. "Cross-border e-commerce platform for commodity automatic pricing model based on deep learning." Electronic Commerce Research 22.1 (2022): 1-20.
14. Dolfen, Paul, Liran Einav, Peter J. Klenow, Benjamin Klopack, Jonathan D. Levin, Larry Levin, et al. "Assessing the gains from e-commerce." American Economic Journal: Macroeconomics 15.1 (2023): 342-370.
15. Bawack, Ransome Epie, Samuel Fosso Wamba, Kevin Daniel André Carillo & Shahriar Akter. "Artificial intelligence in E-Commerce: a bibliometric study and literature review." Electronic markets 32.1 (2022): 297-338.
16. Kethan, M., and Mahabub Basha. "Relationship of Ethical Sales Behaviour with Customer Loyalty, Trust and Commitment: A Study with Special Reference to Retail Store in Mysore City." East Asian Journal of Multidisciplinary Research 1.7 (2022): 1365-1376.
17. Rivaldo, Yandra, Shandrya Victor Kamanda, and Edi Yusman. "The Influence Of Brand Image, Promotion And Trust On Customer Loyalty At Bank BSI Nagoya Batam Branch." Jurnal Mantik 6.2 (2022): 2385-2392.
18. Koay, Kian Yeik, Chee Wei Cheah, and Yi Xuan Chang. "A model of online food delivery service quality, customer satisfaction and customer loyalty: a combination of PLS-SEM and NCA approaches." British Food Journal 124.12 (2022):

4516-4532.
19. Kimura, Makoto. “Customer segment transition through the customer loyalty program.” Asia Pacific journal of marketing and logistics 34.3 (2022): 611-626.
20. Halim, Kevin Yudhaprawira. “Gender classification based on electrocardiogram signals using long short term memory and bidirectional long short term memory.” Jurnal Ilmiah Teknik Elektro Komputer dan Informatika 9.3 (2023): 606-618.
21. Huan, Hai, Zelin Guo, Tingting Cai & Zichen He. “A text classification method based on a convolutional and bidirectional long short-term memory model.” Connection Science 34.1 (2022): 2108-2124.