2023 IEEE 6th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC) | 978-1-6654-6004-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ITNEC56291.2023.10082702

Smartwatch Sales Forecast Based on CNN-LSTM

Lulu Yan
College of E-commerce and Logistics, Beijing Technology and Business University, Beijing, China
13522559892@163.com

***Abstract*—Sales forecast is based on the historical sales volume and related factors to predict future sales volume. An accurate sales forecast can provide a scientific theoretical basis for the overall planning of e-commerce enterprises and upgrading e-commerce products. However, with the rapid development of the Internet, there are more and more factors affecting the change of sales volume, and it becomes more and more challenging to forecast the sales volume of e-commerce products. This paper proposes a sales forecasting method based on CNN-LSTM for sales forecast for smartwatches.Firstly, the dimensionality of the feature data of smartwatches is reduced, and then the processed data is put into the model proposed in this paper to predict sales. At the same time, BP neural network model is used for comparative experiments. The results show that the CNN-LSTM sales forecasting method has a better prediction effect, which is of great significance for the overall planning of e-commerce enterprises.**

Keywords—Sales forecast;smartwatch; CNN-LSTM; feature extraction

I. INTRODUCTION

Intelligent wearable devices are new high-tech wearable devices that intelligently design ordinary wearables based on mature science and technology, such as smartwatches, smart bracelets, smart glasses, and smart clothes. In recent years, smartwatch sales have skyrocketed and are expected to be a key driver of wearable sales growth in the future. However, many factors affect the sales of smartwatches and little literature on their sales prediction. Therefore, this paper proposes a prediction method based on CNN-LSTM to predict the sales of smartwatches.

Firstly, the smartwatch features are extracted, and then the dimensionality of the existing features is reduced by using the low-variance method, Lasso,and principal component analysis (PCA). Finally, put the processed data into the CNN-LSTM model for sales prediction, and the BP neural network model is used for comparison. The experimental results show that the sales forecasting method proposed in this paper can predict sales more accurately.

The main contributions of this study are as follows:

1. In terms of sales forecasting theory, this paper provides a sales forecasting method based on CNN-LSTM, which makes up for the lack of theoretical methods for smartwatch sales forecasting. It provides theoretical support for future sales forecasting of smart wearable devices.

2. At the level of e-commerce enterprises, it provides a scientific basis for the overall planning of e-commerce enterprises and the reduction of operating costs.

3. At the level of intelligent wearable device manufacturers, it provides a direction and scientific basis for upgrading intelligent wearable devices.

The structure of this paper is mainly divided into six parts: The first part describes the background, significance, and contribution of the smartwatch sales forecast. The second part expounds on the current sales forecast and the research in the field of the smartwatch. The third part describes the extraction of smartwatch feature data. The fourth part introduces the model constructed in this paper in detail. The fifth part forecasts the sales volume of the smartwatch and compares the forecast results. The sixth part is the summary and prospect of this paper.

II. RELATED WORK

The sales forecast is to analyze the factors that affect the sales of goods and makes a scientific forecast of future sales scientifically and reasonably [1]. Alekseev K et al. studied seven companies to reduce food waste at grocery retailers, providing essential insights into information systems, forecasting, classification planning, service level definition, replenishment, and waste reduction[2]. According to two typical data features of gasoline demand, Zhang JD et al. constructed a new decomposition ensemble forecasting model driven by trend and cycle characteristics. Experiments show that the model is suitable for sales forecasting in different periods [3]. Warren-vega. W. et al. established a multivariate prediction model for tequila sales. Agave sales are estimated by factors such as the number of plants available for cultivation, total agave production, agave exports, the dollar exchange rate, and annual precipitation[4]. Wu MF and Chen W proposed a PCA-GRNN model to predict the sales of electric vehicles. The factors affecting sales volume were analyzed by PCA dimensionality reduction, and then the data were input into the GRNN model to obtain the future sales volume of electric vehicles[5].

Smartwatch, as the "darling" of smart wearable devices in the new era, has been studied by more and more scholars. Jiang ZF introduced the status quo of the import and export of smartwatches in the post-epidemic era, the difficulties faced, and the specific upgrading road[6]. Wang MM et al. took Apple Watch as an example to deeply analyze the pain points of smartwatch application interaction design, put forward the principles of smartwatch application interaction design, and provided the direction for smartwatch interaction design[7]. Deng WB et al. put forward a method of design evaluation based on the analytic hierarchy process (AHP), the product is decomposed into several stay evaluation indexes and grade evaluation, sorting data and calculating the weight of each index value, finally it is concluded that the scheme of composite scores, in turn, sort, influence factors on children's intelligence watch gets bigger[8]. Jiao HF et al. studied smartwatch user satisfaction based on the ACSI model, and the research results pointed out the direction for smartwatch optimization and service improvement[9]. Yang J et al. mined the online shopping review data of smartwatches and used TF-IDF, N-Gram model, and Word2Vec model to make text vector-quantization and then used machine learning models such as logistic regression, Naive Bayes, support vector machine, and decision tree to classify sentiment propensity. The experimental results show that TF-IDF combined with a support vector machine for text propensity analysis and judgment has the best effect[10].

Although many scholars have studied sales forecasts and smartwatches, there are few studies on sales forecasts for smartwatches. Therefore, this paper proposes a model based on CNN-LSTM to study the sales prediction of smartwatches.

III. THEORETICAL METHOD

This paper aims to study a smartwatch sales forecasting method based on CNN-LSTMand takes the actual sales data of smartwatches on the JD.com's e-commerce platform as an example to make sales forecasting. The main framework is as follows:

Step 1: Data preprocessing, low-variance reduction method, Lasso method,and principal component analysis (PCA) were used to reduce the feature dimension. The feature selection and verification data are all from the real sales data of the JD platform smartwatch, and the specific content will be elaborated on in Sections III.

Step 2: Build the model, input the features screened in Step 1 into different prediction models, compare the performance of different models, and select the optimal model. The specific content will be shown in Sections IV and V.

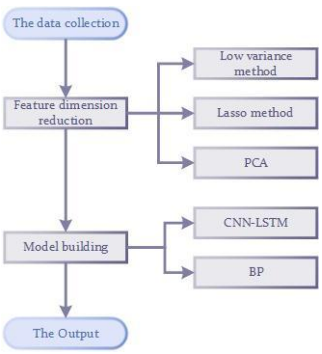

Fig. 1. Experimental structure diagram

A. Low-variance treatment

Low-variance treatment is the removal of low-variance features. A low-variance indicates that the features are uniform and the discrimination is low. The low-variance removal method can be used as a primary feature screening method. Generally, the threshold can be set larger to delete features too uniform in value. In this experiment, the threshold is set as 0.8, and characteristics retained after low-variance treatment are shown in TABLE I.

TABLE I. EIGENVALUES AFTER LOW-VARIANCE TREATMENT

The attributes	*Characteristics*
Appearance attributes	dial shape;screen type;wrist strap material;wrist strap color
Functional properties	automatic adjustment of brightness;battery life;blood oxygen monitoring;exercise mode;AI voice assistant;real-time micro-chat; emergency help;installation software; call type;flashlight
Other attributes	product praise rate;brand; average price;time to market

B. Lasso method processing

Feature selection is based on the loss function(the optimization goal) to join the penalty term, which will make the training to solve the parameters in the process of considering the coefficient of size by setting the reduction coefficient (punish coefficient), which will make the characteristics of a smaller effect coefficient of attenuation to 0, only keep important characteristics, the process of the Lasso algorithms is the model coefficient is less than the sum of the absolute value of constant conditions, To minimize the residual sum of squares. In this experiment, 0/1 coding was performed for features with values of two categories (whether there is blood oxygen monitoring, whether there is emergency help, etc.), and one-hot coding was performed for features with values greater than two categories (call type, wristband color, etc.). All categories were retained, and a total of 29 features were obtained. Characteristics retained after lasso treatment are shown in TABLE II.

TABLE II. EIGENVALUES AFTER LASSO TREATMENT

The attributes	*Characteristics*
Appearance attributes	dial shape _ round;wrist strap material _ fluorine glue;wrist strap color _ white
Functional properties	AI voice assistant; flashlight;call type _ not supported
Other attributes	product praise rate;average price;time to market;brand _

	huawei;brand _ honor

C. PCA feature dimension reduction

Principal Components Analysis (PCA) the purpose of PCA is to reduce the dimensionality of features, use orthogonal transformation to linearly transform the eigenvalues of a series of variables that may be correlated, and project them into the values of a series of linearly uncorrelated variables, namely Principal components. PCA maps the data to a lower dimensional space, which can reduce the amount of data computation for the next prediction model. In this experiment, a total of four principal components are generated after PCA dimensionality reduction, whose calculation formula is as follows:

TABLE III. EIGENVALUES AFTER LASSO TREATMENT

The attributes	*Characteristics*
X_1	product praise rate
X_2	AI voice assistant;
X_3	flashlight
X_4	average price
X_5	time to market
X_6	dial shape _ round
X_7	wrist strap material _ fluorine glue
X_8	wrist strap color _ white
X_9	call Type _ Not supported
X_{10}	brand _ huawei
X_{11}	brand_honor

$F1=-0.352*X_1+0.083*X_2+0.396*X_3-0.340*X_4-0.396*X_5+0.396*X_6+0.315*X_7-0.218*X_8+0.189*X_9+0.083*X_{10}+0.306*X_{11}$

$F2=+0.143*X_1+0.551*X_2+0.100*X_3+0.239*X_4-0.100*X_5+0.100*X_6+0.230*X_7-0.079*X_8-0.303*X_9+0.551*X_{10}-0.366*X_{11}$

$F3=-0.042*X_1-0.056*X_2+0.021*X_3-0.275*X_4-0.021*X_5+0.021*X_6+0.343*X_7+0.772*X_8-0.443*X_9-0.056*X_{10}+0.067*X_{11}$

$F4=+0.022*X_1+0.225*X_2+0.083*X_3+0.011*X_4-0.083*X_5+0.083*X_6-0.304*X_7+0.563*X_8+0.675*X_9+0.225*X_{10}-0.110*X_{11}$

IV. CNN - LSTM MODEL

A. The CNN

The word "neuron" first appeared in biology, and researchers created neural networks that mimic how neurons respond to stimuli. A convolutional neural network (CNN) is more like a human visual neuron, which can receive a large amount of information, focus on locally important information, and finally synthesize all local essential information for judgment and processing. Unlike traditional neural networks, CNN describes an end-to-end mapping relationship, which can significantly reduce the complexity of data and the number of parameters for model training while maintaining the original features through a series of operations such as convolution, pooling,and complete connection. CNN has the ability to learn complex mapping and high-dimensional space. The main structure includes a convolutional layer, pooling layer, activation function,and fully connected layer.

B. LSTM

A recurrent neural network (RNN) is a neural network with "memory" ability. It not only considers the influence of the input value on the output value at the current stage but also considers the influence of the output value at the previous stage on the output value at the current stage, realizes the connection between each hidden layer, and can capture the relationship between the model time series. But because its weight to nod at each time step is the same, along with the increase in data, multiplied by the weight matrix, the weighting coefficient of exponential growth easily lead to " gradient disappearance " and "gradient explosion", make the distance prediction point far data can not be effectively used, not to solve the problem of the "long relied on" the neural network.

The Long short-term memory network (LSTM) model is modified based on RNN, which can well solve the problems of "gradient disappearance" and "gradient explosion" caused by RNN "unable to rely on for a long time." The main improvement of LSTM is adding a memory cell state and "gate" structure. The memory cell is used to record historical information, which is constantly updated and transmitted in the cell state. The "gate" structure is used to manage the transmission of information. The "gate" structure comprises an input, forget, and output gate. It does not store information and only manages the transfer of information through the sigmoid function. The structure of the LSTM model is shown in Fig. 2.

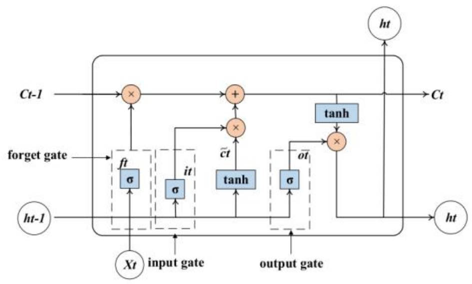

Fig. 2. LSTM structure

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i * [h_{t-1}, X_t] + b_i) \quad (2)$$

$$\widehat{C}_t = tanh(W_c * [h_{t-1}, X_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \widehat{C}_t \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \quad (5)$$

$$h_t = o_t * tanh(C_t) \quad (6)$$

In the above equation, h_{t-1} is the output value of the previous state, X_t is the input value of the current moment, σ and tanh all are activation functions, b_f, $b_i, b_c\ and\ b_o$ all are bias terms, $W_f, W_i, W_c\ and\ W_o$ all are weight matrices.

C. CNN - LSTM

This paper proposes a fusion model based on CNN and LSTM, namely CNN-LSTM model. The features are input into the CNN model for convolution and other operations to extract meaningful information from the features to reduce the amount of calculation for the model calculation in the following steps.Then the convolution data is input into the LSTM model for sales prediction. The model structure is shown in Fig. 3:

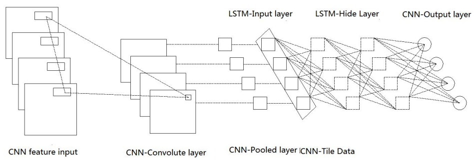

Fig. 3. Structure diagram of CNN-LSTM model

V. EXAMPLE VERIFICATION

A. Experimental Environment

In order to verify the effectiveness of the combined model the CNN-LSTM for smartwatch sales prediction, CNN-LSTM model and BP neural network are constructed on the Keras framework, which supports Python language.

B. Data Sources

This data comes from Jingshu Data Company's monitoring of JD's smartwatch sales volume, and the feature data comes from JD's webpage information extraction.

C. Evaluation Index

MAE (Mean Absolute Error) is the average value of Absolute Error, which can better reflect the actual situation of the predicted value Error.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\widehat{y_i} - y_i| \quad (7)$$

MSE (Mean Squared Error) is a convenient method to measure the average error, which can evaluate the error between the actual value and the predicted value. Root mean square error is the arithmetic square root of mean square error.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2 \quad (8)$$

D. Experimental Results

In order to verify the superiority of the CNN-LSTM model proposed in this paper, this study selects 101 weeks of actual sales data of a smartwatch on the JD e-commerce platform from 27 weeks in 2020 to 23 weeks in 2022 for experiments. It uses the traditional BP neural network as the comparison model. The MAE and MSE of the two models are respectively compared, and the results are shown in the figure below. The evaluation indexes of the CNN-LSTM model are lower than those of the BP neural network model, and the accuracy of the CNN-LSTM model in this experiment is 24% higher than that of the BP neural network model,which accuracy rate can reach 96.2%. It proves that the CNN-LSTM model proposed in this paper has a better effect on smartwatch sales prediction.

TABLE IV. EXPERIMENTAL RESULT

Moudels	*MAE*	*MSE*
CNN-LSTM	0.0233	0.0008
BP	0.1822	0.0386

VI. CONCLUSION AND PROSPECT

To sum up, based on the fusion of the CNN and LSTM models, this study constructs the CNN-LSTM model to extract the features that affect the sales of smartwatches and then put them into the CNN-LSTM model for sales prediction. The experimental results show that the CNN-LSTM model has higher accuracy than the traditional BP neural network. The prediction model constructed in this study performs well in MAE and MSE. Through the actual smartwatch sales data of the Jingdong e-commerce platform, the experiment shows that the CNN-LSTM model constructed in this paper can predict the sales trend of smartwatches more accurately and can be used for actual sales prediction. However, due to the rapid replacement of smartwatches, a large number of competing products, and a large number of online evaluation information, future research should take competitive product information, online public opinion, and other factors into account to make more accurate sales predictions.

REFERENCES

[1] Alekseev K and Seixas J M, “Forecasting the air transport demand for passengers with neural modelling,”IEEE Computer Society.2002

[2] Zhang JD and Zhao JL,”Trend- and Periodicity-Trait-Driven Gasoline Demand Forecasting,”Energies.15(10), 3553,2022

[3] Lena R and Alexander H,”Reducing Food Waste at Retail Stores—An Explorative Study,”Sustainability.14(5), 2494,2022

[4] Warren-Vega WM,Aguilar-Hernandez DE,Zarate-Guzman AI,Campos-Rodriguez A and Romero-Cano LA,”Development of a Predictive Model for Agave Prices Employing Environmental, Economic, and Social Factors: Towards a Planned Supply Chain for Agave-Tequila Industry,”.Foods.11(8), 1138,2022

[5] Wu MF and Chen W, “Forecast of Electric Vehicle Sales in the World and China Based on PCA-GRNN,”Sustainability. 14(4), 2206,2022

[6] Jiang ZF,”China's smart watch export faces difficulties and the way to upgrade,”Foreign Economic and Trade Practice,2021(11):59-62.

[7] Wang MM and Zeng FG,”Research on Application Interaction Design of Smart Watch Based on Application Scenarios -- Taking

Apple Watch as an Example,"Art and Design (Theory).2(08):98-100,2019

[8] Deng WB and Zhu HX,"Research on design evaluation of children's smart watch based on analytic hierarchy process,"Packaging engineering.39(08):121-125,2018

[9] Jiao HF and He M,"Research on user satisfaction of wearable devices based on ACSI model -- taking smart watch as an example,"Modernization of shopping malls.2021(19):1-5.

[10] Yang J and Zhou J,"Research on Emotional Tendency of Online Shopping Comments on Smart Watch,"Industrial Control Computer.34(04):97-99,2021