Three Basic Deep Learning Network Structures for E-Commerce Logistics Demand Forecasting

Shengjie Ke[1]

[1] *Fuzhou University of International Studies and Trade, Fuzhou, Fujian, China*
[1]*Corresponding author. Email: keshengjie@fzfu.edu.cn*

ABSTRACT

Logistics demand forecasting is an important business in the e-commerce field, involving various types of data with complex features; deep learning, as a powerful machine learning method, can automatically extract high-level feature representations from data, improving prediction performance. This article introduces three basic deep learning network structures suitable for e-commerce logistics demand forecasting, namely the spatial feature extraction network structure based on convolutional neural network (CNN), the temporal feature extraction network structure based on recurrent neural network (RNN), and the multi-dimensional feature fusion network structure based on attention mechanism. This article also proposes methods to evaluate the prediction performance of these network structures, including data sets, evaluation indicators, benchmark methods and evaluation processes.

Keywords: *E-commerce, Logistics demand forecasting, Deep learning.*

1. INTRODUCTION

E-commerce logistics demand forecasting refers to using mathematical models or machine learning methods based on historical sales data, user behavior data, product attribute data, etc. of e-commerce platforms to predict the logistics demand in a future period, including order quantity, product type, distribution area, etc. E-commerce logistics demand forecasting is of great significance for optimizing logistics resource allocation, improving logistics efficiency, reducing logistics costs, and improving user satisfaction. However, e-commerce logistics demand forecasting also faces many challenges, such as the high dimensionality, sparsity, nonlinearity, and dynamics of data. Traditional mathematical models or machine learning methods often have difficulty in effectively dealing with these complex data features, resulting in low prediction accuracy. Deep learning, as a powerful machine learning method, can automatically extract high-level feature representations from raw data through a multi-layer neural network structure, thereby improving prediction performance. However, the design of deep learning network structure is a challenging work that requires considering factors such as network complexity, parameter quantity, training time, generalization ability, etc. Different network structures may have different adaptability to different data features and prediction tasks.

This article aims to introduce three basic deep learning network structures suitable for e-commerce logistics demand forecasting. First, it explains the design ideas, principles and application processes of each network structure respectively. Then it details the various components and functions of these three network structures. At the same time, it proposes datasets that can be used to evaluate the predictive performance of these network structures, suitable evaluation metrics, benchmark prediction methods for comparison, and the evaluation process.

The three basic deep learning networks proposed are:

- (1) Spatial feature extraction network structure based on convolutional neural network (CNN)[1]
- (2) Temporal feature extraction network structure based on recurrent neural network (RNN)[2]

- (3) Multi-dimensional feature fusion network structure based on attention mechanism (Attention)[3]

2. RELATED RESEARCH

In recent two years many scholars have published research results on building deep learning networks to predict e-commerce logistics demand. Neelakandan S et al. (2023) adopted a continuous stochastic fractal search (SFS) method to optimize the parameters of the deep learning improved neural network (DLMNN), achieving performance improvement of the DLMNN model on multiple product categories[4]. Petroșanu D M et al. (2022) proposed a dynamic directed acyclic graph neural network (DAGNN) deep learning architecture that can make fine-grained predictions of daily sales revenue for each product category up to three months in advance[5]. Gandhi A et al. (2022) proposed a graph neural network and deep learning-based multivariate correlated time series forecasting framework for predicting demand for different sellers and products on Amazon's online marketplace that can capture homogenous and heterogeneous correlations between sellers and products as well as consider the impact of factors such as competition out-of-stock new products etc.[6]. Wang Yiwen Wang Weili (2022) proposed a combination model LSTM-RELM of long short-term memory network (LSTM) and regularized extreme learning machine (RELM) for predicting daily total amount of e-commerce product transactions[7].

However the deep learning network structures constructed by the above research are relatively complex and it is difficult to reproduce the corresponding prediction effects in practical applications. The following will introduce three easy-to-implement basic deep learning network structures.

3. THREE BASIC DEEP LEARNING NETWORK STRUCTURES SUITABLE FOR E-COMMERCE LOGISTICS DEMAND FORECASTING

E-commerce logistics demand forecasting is a method that uses historical data and other relevant factors to predict future logistics demand which can help e-commerce enterprises optimize inventory management reduce operating costs and improve customer satisfaction. Deep learning is a machine learning technique based on artificial neural networks that can automatically learn features and rules from large amounts of data and is suitable for complex and nonlinear problems. Here are three deep learning network structures suitable for e-commerce logistics demand forecasting.

3.1 Spatial Feature Extraction Network Based on Convolutional Neural Network (CNN)

The spatial feature extraction network based on convolutional neural network (CNN) can extract the spatial distribution features of logistics demand from historical data and use these features to predict future demand.

3.1.1 Design Ideas and Principles

The design idea of the network structure is to treat the logistics demand as a two-dimensional matrix, one of the dimensions is time, the other dimension can be the geographical area, each element represents the logistics demand of the region at a certain moment, and then use CNN to convolve the two-dimensional matrix to extract the spatial characteristics of different scales and directions. The principle of this network model is that the convolution kernel of CNN can capture local correlations and translational invariance in two-dimensional matrices, thereby effectively learning spatial features.

3.1.2 Application Process

The application process of the network structure first converts the historical data into a series of two-dimensional matrices, and then input these two-dimensional matrices into the CNN, and obtains the final two-dimensional matrix of spatial features after multi-layer convolution and pooling operations, and finally uses the fully connected layer or other regression models to predict the two-dimensional matrix of spatial features and outputs a two-dimensional matrix containing future requirements.

3.1.3 Individual Components and Functions

Convolutional layer: Multiple convolution checks are used to perform sliding window operations on the input two-dimensional matrix to generate multiple feature two-dimensional matrices, each feature two-dimensional matrix reflects the spatial characteristics of the input two-dimensional matrix at a certain scale and direction.

Activation function: Nonlinear transformation of the output of the convolutional layer increases the expressiveness and generalization ability of the network model.

Pooling layer: Down sampling the output of the convolutional layer reduces the number of parameters and computation while retaining important spatial characteristics.

Fully connected layer or other regression model: This model performs a linear or other nonlinear transformation of the output of the pooled layer to map the two-dimensional matrix of spatial features to a two-dimensional matrix for future needs.

3.2 Network Structure of Time Series Feature Extraction Based on Recurrent Neural Network (RNN)

To understand the working principle and components of the network structure of time series feature extraction based on recurrent neural network (RNN), the author will discuss this in two parts. First, the author will explore the design concept and principle, focusing on how this network utilizes the properties of RNNs to process time-series data and predict future logistics demands. Secondly, the author will provide a detailed description of each component of the network and their functions, including RNN units and fully connected layers or other types of regression models.

3.2.1 Design Ideas and Principles

The network structure is designed to use the memory ability and dynamism of RNNs to extract useful features from time series data, and then use these features to predict future logistics needs. The principle of the network structure is to divide the input time series data (such as sales volume, price, seasonality, etc.) into several time steps in chronological order, and then input the data of each time step into an RNN unit, which will output a new hidden state according to the current input and the previous hidden state, which contains the characteristic information of the current time step. In this way, after multiple time steps, the hidden state of the last time step contains the characteristic information of the entire time series data, and then this hidden state is entered into a fully connected layer or other regression model to obtain the predicted value of future logistics demand.

3.2.2 Individual Components and Functions

RNN unit: It is the core part of the network structure, responsible for extracting feature information from time series data. RNN units can be of different types, such as simple RNNs, long short-term memory (LSTMs), or gated recurrent units (GRUs), all of which have similar structures, but use different activation functions and gating mechanisms to process inputs and outputs to accommodate data with different characteristics. The input of the RNN unit is the data of the current time step and the hidden state of the previous time step, and the output is the hidden state of the current time step and the optional output value.

Fully connected layer or other regression model: It is the output part of the network structure, which is responsible for predicting future logistics requirements based on the extracted feature information. Fully connected layers or other regression models can have different types, such as linear regression, polynomial regression, or neural network regression, all of which have similar structures but differ in fitting complexity and generalization ability. The input of the fully connected layer or other regression model is the hidden state of the last time step, and the output is the predicted value of future logistics demand.

3.3 Multi-Dimensional Feature Fusion Network Structure Based on Attention Mechanism

In this section, we will delve into the multi-dimensional feature fusion network structure implemented using the Attention mechanism. First, we will introduce the design concept and basic principles of this network structure, namely how the Attention mechanism captures the relevance and importance between different dimensional features to enhance the accuracy and robustness of predictions. Following that, we will provide a detailed description of each component of the network and their functions, including the fully connected layer, the self-attention layer, the weighted sum layer, and the output layer.

3.3.1 Design Ideas and Principles

The design idea of the network structure is to use the attention mechanism to capture the correlation and importance between features in different dimensions, to improve the accuracy and robustness of prediction. The principle of the network structure is to map the input multidimensional feature vector to a low-dimensional space through a fully connected layer, and then calculate the weight of each feature vector through a self-attention layer, and then obtain a fused feature vector through a weighted summation layer, and finally obtain the prediction result through an output layer.

3.3.2 Individual Components and Functions

Fully connected layer (FC): The function of this layer is to map the input multidimensional feature vector to a low-dimensional space, reduce the dimension of the feature, and reduce the computational complexity.

Self-attention layer: The role of this layer is to calculate the weight of each feature vector, reflecting its relevance and importance in prediction. The layer first multiplies each feature vector by three different weight matrices to obtain the query vector (Query), key vector (Key) and value vector (Value). Then by calculating the dot product between the query vector and the key vector, the Attention Score is obtained. Then through a scaling factor and a SoftMax function, the attention weight is obtained. Finally, by multiplying the attention weight and the value vector, the Output Vector is obtained.

Weighted sum layer: The function of this layer is to add multiple output vectors output from the attention layer to obtain a fused feature vector that reflects the overall information of the input features.

Output layer: The function of this layer is to pass the fusion feature vectors output by the weighted sum layer through an activation function to obtain the prediction result.

4. PREDICTIVE PERFORMANCE EVALUATION

To evaluate the prediction performance of the three deep learning network structures proposed in this paper, it is necessary to use a dataset containing e-commerce logistics requirements, select appropriate evaluation indicators and benchmark prediction methods for comparison, and design a reasonable evaluation process.

4.1 Datasets

The dataset containing e-commerce logistics requirements can be selected from a public dataset called "Online Retail II" from the UC Irvine Machine Learning Repository. The dataset, provided by an online retailer in a non-physical store in the UK, recorded all transaction data between December 1, 2009, and December 9, 2011. The company mainly sells unique gifts for various occasions, and many customers are wholesalers. The dataset contains 8 attributes, namely order number, product code, product name, quantity, order date, unit price, customer number and country. This dataset is suitable for evaluating the predictive performance of deep learning network structures for e-commerce logistics demand forecasting because it has the following characteristics:

- It is a multivariate, time-series, text-containing data set that can reflect the change of customers' purchasing behavior and preferences over time.
- It contains many actual transaction data, which can provide sufficient sample size and diversity, and has a certain degree of noise and missing values, which can investigate the generalization ability and robustness of the model.
- Involving different product categories, prices and countries, the performance and adaptability of the model when processing data of different dimensions and scales can be examined.

4.2 Evaluation Indicators

Metrics that can be used to measure the difference between the predicted value and the true value include mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and so on.

4.3 Baseline Forecasting Methodology

Here are three baseline forecasting methods that people can use for comparison.

- Linear Regression (LR): A simple linear model that assumes a linear relationship between e-commerce logistics demand and input characteristics.

- Autoregressive Integrated Moving Average (ARIMA): A commonly used time series analysis model that uses historical data to predict future trends and seasonal changes.
- Facebook forecasting tool Prophet: A forecasting tool that automatically handles trend, seasonality, and holiday effects in time series.

4.4 Evaluation Process

First, the dataset is divided into training set, validation set, and test set according to 80%, 10%, and 10% ratios. Secondly, the deep learning network structure based on CNN, RNN and Attention is constructed respectively, and the training set is used for model training. Third, the validation set is used for model tuning and the optimal hyperparameters are selected. Finally, the test set is used for model evaluation.

According to the above evaluation process, the three deep learning network structures and three benchmark forecasting methods can be experimentally compared, the advantages, disadvantages and applicable scenarios of each method in e-commerce logistics demand forecasting can be analyzed, and the performance of each method on different evaluation indicators and the performance of each method on different time scales and forecasting tasks can be analyzed.

- Overall prediction performance: comparing the prediction performance of each method on the entire test set and analyzing the overall strengths and weaknesses of each method.
- Prediction performance over time: comparing the prediction performance of each method in different time periods and analyzing the ability of each method to capture the trend and seasonality changes in the time series.
- Prediction performance by product category: comparing the prediction performance of each method on different product categories and analyzing the adaptability of each method to different product features and demand patterns.
- Prediction performance by country: comparing the prediction performance of each method on different countries and analyzing the adaptability of each method to different country markets and demand differences.

5. CONCLUSION

Aiming at the problem of e-commerce logistics demand forecasting, which has theoretical research significance and practical application value, this paper proposes three easy-to-implement basic deep learning network structures suitable for this problem, namely the spatial feature extraction network structure based on CNN, the time series feature extraction network structure based on RNN, and the multi-dimensional feature fusion network structure based on Attention. The components and functions of the three network structures are described in detail, and the datasets, suitable evaluation indicators, benchmark prediction methods for comparison, and evaluation processes are proposed to evaluate the prediction performance of the model.

ACKNOWLEDGMENTS

Funding Project: This article is one of the research results of the 2021 Fujian Province Young and Middle-aged Teachers Education Research Project (Science and Technology Category) "Research on Deep Learning Network Structures Suitable for E-commerce Logistics Demand Forecasting" (Project Number: JAT210533).

REFERENCES

[1] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position [J]. Biological cybernetics, 1980, 36(4): 193-202.

[2] Lipton Z C, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning [J]. arXiv preprint arXiv: 1506.00019, 2015.

[3] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint arXiv: 1409.0473, 2014.

[4] Neelakandan S, Prakash V, Pranav Kumar M S, et al. Forecasting of E-Commerce System for Sale Prediction Using Deep Learning Modified Neural Networks [C]//2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC). IEEE, 2023: 1-5.

[5] Petroșanu D M, P îjan A, Căruţaşu G, et al. E-Commerce Sales Revenues Forecasting by Means of Dynamically Designing, Developing and Validating a Directed Acyclic Graph (DAG) Network for Deep Learning [J]. Electronics, 2022, 11(18): 2940.

[6] Gandhi A, Aakanksha, Kaveri S, et al. Spatio-temporal multi-graph networks for demand forecasting in online marketplaces [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer International Publishing, 2021: 187-203.

[7] WANG Yiwen, WANG Weili. Research on GMV Prediction of E-commerce Based on LSTM-RELM Combination Model [J]. Journal of Computer Engineering & Applications, 2023, 59(10):321-327.