Inventory prediction based on CNN-LSTM Model

He Zhang[2, *, #], Shaokai Tian[1, #], Jianfeng Qin[2, #], Zhengtian Yu[2, #], Huabo Guo[1, #], Xuefeng Jia[2, #]

[1]School of Mechanical Engineering, Tianjin University of Science & Technology, Tianjin, 300222, China

[2]School of Electronic Information and Automation, Tianjin University of Science & Technology, Tianjin, 300222, China

[*] Corresponding Author Email: zhanghetust@163.com

[#]These authors contributed equally.

Abstract. Demand forecasting and inventory optimization are the core issues of e-commerce supply chain management, which are crucial for ensuring timely delivery of goods, reducing inventory costs, and improving inventory turnover efficiency. This paper is based on Python's merge function to merge table data and transcode text data. Then, by drawing a q-q diagram and K-S test to determine the distribution of the data columns, it was found that the shipment volume data follows a normal distribution. Adopting 3σ The principle is to determine outliers and manually determine that the two data with larger values in the edge values are considered outliers. Replace the above outliers with missing values and use Newton interpolation for linear filling to obtain the standard dataset after data preprocessing. We establish a prediction model based on attention mechanism for prediction, and ultimately obtain more accurate prediction results.

Keywords: K-S test, ARIMA model, Similarity algorithm, Clustering algorithm.

1. Introduction

In the field of e-commerce retail, supply chain management is a crucial task, covering multiple aspects such as inventory management, fulfillment speed, and cost-effectiveness[1]. Among them, demand forecasting plays a core role in supply chain management, helping merchants and e-commerce platforms manage inventory more reasonably to ensure timely customer demand and avoid resource waste caused by excessive inventory backlog.

There have been much similar work about predicting indexes by means of CNN models. Sun et.al [2] constructed a CNN-LSTM model for predicting the Shanghai Stock indexes which is helpful for us to transform. Moreover, Song et.al tended to predict the heating system by a hybrid CNN-LSTM model[3]. The work is similar and enlighten our work.

We firstly Predict the demand for goods from various merchants in each warehouse from May 16, 2023 to May 30, 2023, and evaluate the predictive performance of your model. Find similar sequences and complete the predicted values for these dimensions from May 16, 2023 to May 30, 2023. Refer to the provided data and provide the predicted values from June 1st, 2023 to June 20th, 2023.

2. Preliminary

2.1. Assumptions

1.The impact of external factors (such as macroeconomic conditions, policy changes, etc.) on the predicted results is ignored.

2.The demand for goods may be seasonal.

3.Specific promotional activities, such as 618 and Double Eleven, may have a significant impact on sales.

4.The geographical location of the warehouse may affect the delivery time to specific areas, thereby affecting the demand in certain regions.

5.Returned and exchanged goods may not be immediately restocked and sold, which may affect overall demand and shipment volume.

2.2. K-Means Algorithms

The goal of the K-Means algorithm[4] is to minimize the sum of squared errors within the "cluster" after clustering, as the smaller the E-value, the higher the similarity of the samples within the "cluster". However, this minimization process is an NP hard problem, and without an effective algorithm, one can only traverse all possible combinations ("clusters"). When the sample size is large, it is a waste of resources or even without a solution. So consider using greedy algorithms, which seek approximate solutions through iterative optimization. As shown in Equation 1.

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} \| x - \mu_i \|_2^2 \quad (1)$$

2.3. XGBoost Algorithms

The overall idea of XGBoost[5] is to directly add the loss function and regularization term to form a global loss function. The second derivative of this loss function is obtained to obtain the final obj, and a score is calculated using obj. The smaller the score, the better. Finally, the score calculated using obj determines the structure of the tree and the score of the entire strong learner. So XGBoost is not achieved by fitting residuals, but by directly calculating the obj function to obtain the tree structure.

The objective function consists of two parts. The first part is the model error, which is the difference between the true real value and the predicted value of the sample. The second part is the structural error of the model, which is the regularization term. By using hyperparameters to multiply the number and value of nodes, the complexity of the model is limited.

2.4. Notations

The important notations are shown in Table 1.

Table 1. Symbols notation

Symbols	Description
Δy_t	Order difference
p	Lag orders
$f[x_0, x_1, \cdots, x_n]$	Feature vector
x_i	i-th missing item
n	Number of missing items
$Cov(x, y)$	Sample covariance
S_x	Sample standard deviation
r_{xy}	Sample Pearson correlation coefficient
y_t, y_{t-1}	the value of the variable at time t and time t-1
ξ	normal distribution
ε_t	Individual variable coefficients
$\beta_1, \beta_2, \cdots, \beta_4$	White noise sequence
y_{pre}	Prediction value
y_{real}	Real value
$\text{forecast}(t)$	Estimate
$a\mathit{ctual}(t)$	True value

3. Experiments

3.1. Data preprocessing

The data used in this paper comes from the National Competition of Mathematics Modeling. For the given data, there are a large number of datasets that cannot be directly modeled. Subsequent models usually require numerical input, which includes classification or textual data (such as product classification, warehouse area, etc.), and these non numerical data need to be converted into numerical format. Therefore, data transcoding is necessary to fix inconsistencies.

In the process of determining outliers, we first need to ensure that the dataset conforms to a normal distribution. To achieve this goal, we will use MATLAB to draw P-P diagrams and Kolmogorov Smirnov tests on the collected dataset to determine its distribution pattern. For data that does not conform to a normal distribution, we will use manual judgment to process it.

After confirming that the dataset conforms to a normal distribution, we will define outliers based on principles and use this standard to filter out outliers. In order to prevent false positives, we will also make manual judgments. For data that cannot pass the normal distribution test, we will directly make manual judgments. For the missing value, if it is no longer used directly, it will definitely have a certain impact on the results. Therefore, interpolation filling is used for processing here. For the missing data, which contains several nodes with missing outliers, this is the data we need to study the subsequent problem. Therefore, for this part of the missing values, we use Newton interpolation[4] to supplement it.

Some of the results are as follows Table 2.

Table 2. Some of the results

Sale Code	Production Code	Repository Code	Date	Need	Classification
seller_32	product_1091	wh_1	2022/12/12	2606	Construction material
seller_32	Product_1091	wh_1	2022/12/10	3818	Construction material

3.2. Model establishment

To process the data to adapt to the LSTM model[6], the following steps:Data Merge, Feature Program, Time serialization,etc are necessary.

It can be seen that as the number of training rounds increases, both the training error and validation error continue to decrease, indicating that the model gradually converges during the training process and there is no overfitting phenomenon. Use test sets for prediction, obtain shipment volume prediction results, and destandardize.

3.3. Combination CNN and LSTM

The results of combined CNN and LSTM model are shown in Figure 1.

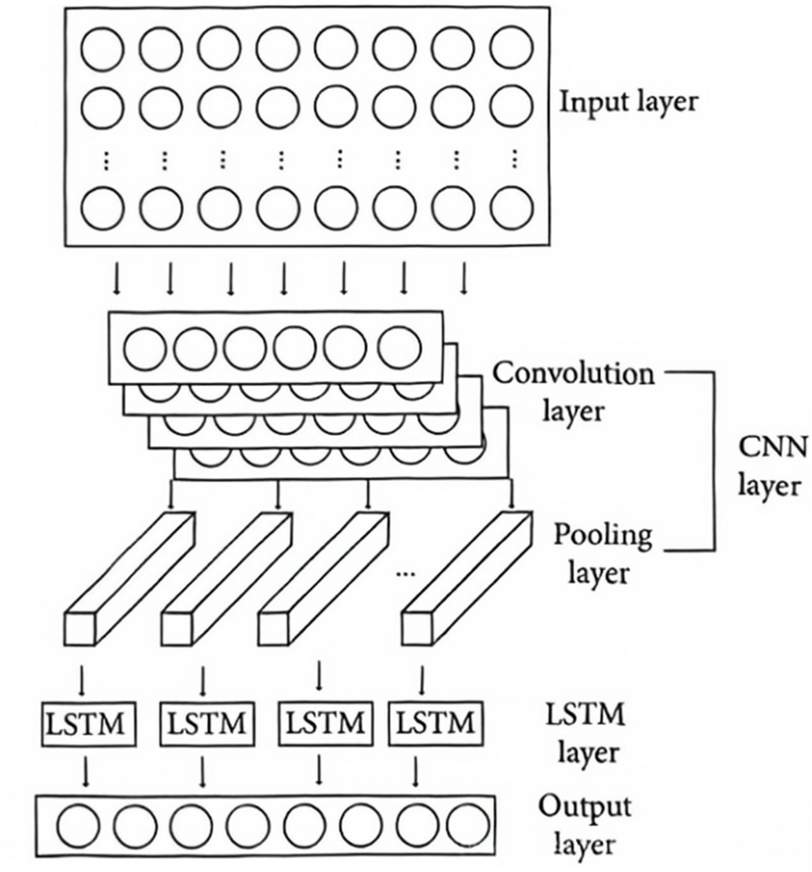

Figure 1. CNN and LSTM combination model

MLP[7] consists of an input layer, a hidden layer, and an output layer. The input layer receives input data, the hidden layer learns feature representations, and the output layer produces the final prediction result. Each neuron in the hidden layer and output layer has an activation function for introducing nonlinear mapping.

3.4. Result

In order to test the prediction results of the prediction model, this paper uses root mean square error, average absolute percentage error, and determination coefficient as evaluation indicators for establishing a time based economic prediction model in this paper.

Cross validation[8] usually divides the dataset into *k* subsets, with *k*-1 subset as the training set and the remaining 1 subset as the testing set. This allows for *k* training and testing sessions, and ultimately averages the k test results to obtain the performance indicators of the model. By repeating this process multiple times, more accurate performance indicators can be obtained. For the stability of the model, this article will use cross validation method to evaluate the CNN LSTM prediction model.

Table 3 shows the comparison of Results of Four Prediction Models.

Table 3. The rank of each countries

Index	Root mean square error	Mean Absolute Percent Error	Coefficient of Determination
CNN-LSTM	4.1332	3.6052	0.9652

It can be seen that the weighted prediction model has the best results for each indicator. The root mean square error is the smallest, expressed as the sum of squared errors between the predicted value and the true value. The higher the numerical value, the higher the accuracy, indicating that the weighted prediction model has the best results; The average absolute percentage error is relatively small, indicating that the coefficient is closest to 1, indicating that the closer the sum of squares of the error is to the sum of squares, the better the prediction effect.

In summary, the three prediction models established using LSTM prediction model, time series model, and multiple linear regression prediction have good prediction accuracy for the coefficients solved by the optimized model constructed with the minimum error of the results.

4. Discussion

From the detection, it can be seen that the model can achieve purchase volume prediction to a certain extent and has good experimental results.

The final predicted results are as Table 4.

Table 4. Predict partial results

seller_no	product_no	warehouse_no	date	forecast_qty
seller_10	product_1915	wh_1	2023/5/16	5
seller_10	product_1917	wh_1	2023/5/16	1
seller_10	product_1914	wh_1	2023/5/16	1
seller_10	product_1916	wh_1	2023/5/16	0

To avoid running for too long, it is common to set a maximum iteration round or a minimum adjustment amplitude threshold. Here, the time part method is used to determine the K value. The core idea of the hand time method is that as the number of clusters *k* increases, the sample division will become more refined, and the degree of clustering of each cluster will gradually increase. Therefore, the sum of error squares and SSE will naturally decrease. Moreover, when *k* is less than the number of nearest clusters, the increase in k will significantly increase the aggregation degree of each cluster, resulting in a significant decrease in SSE. However, when *k* reaches the optimal number of clusters, the aggregation degree obtained by increasing *k* will quickly decrease, resulting in a sharp decrease in SSE's decline[9,10]. As the value of k continues to increase, it tends to flatten out, which means that the relationship between SSE and *k* is in the shape of a hand clock, And the *k* value corresponding to this time part is the optimal number of clusters for the data.

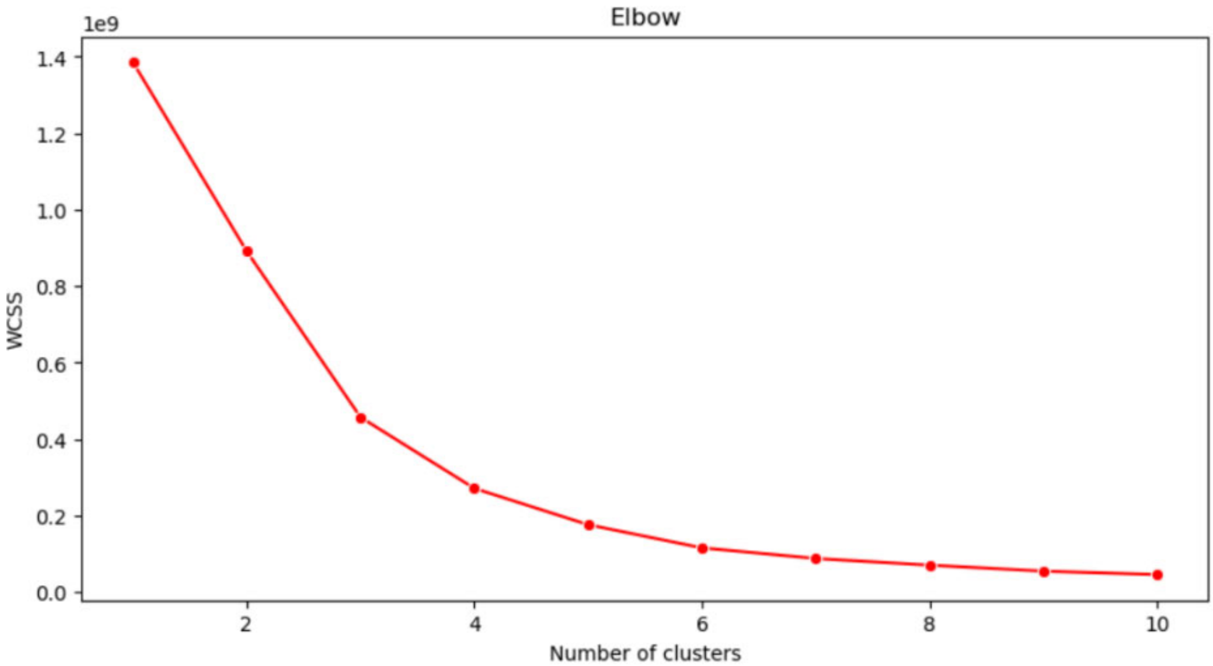

Figure 2. Ranking of host cities of the Summer Olympics

From the Figure 2, it can be seen that selecting K=3.

The clustering results are as Figure 3 and Figure 4.

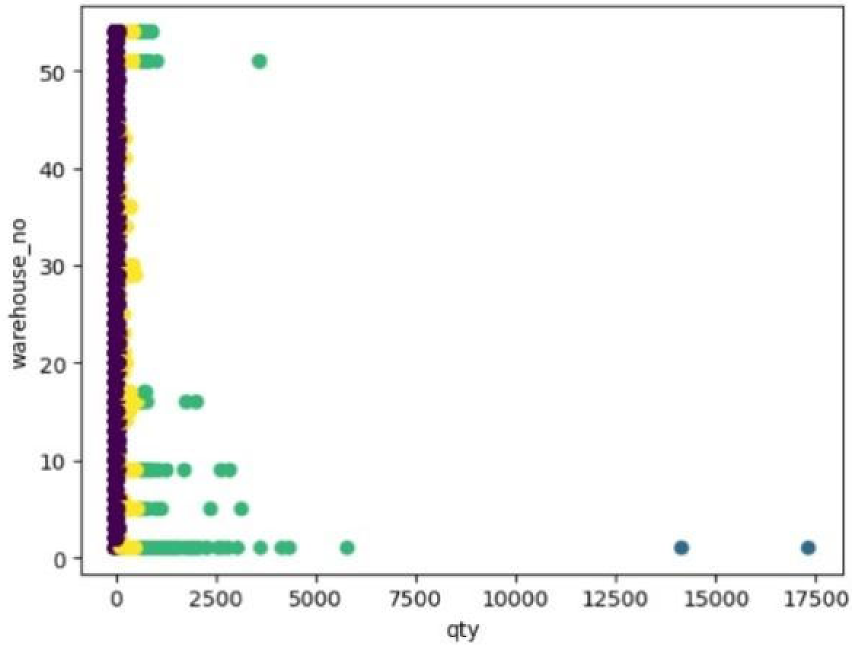

Figure 3. Clustering results(1)

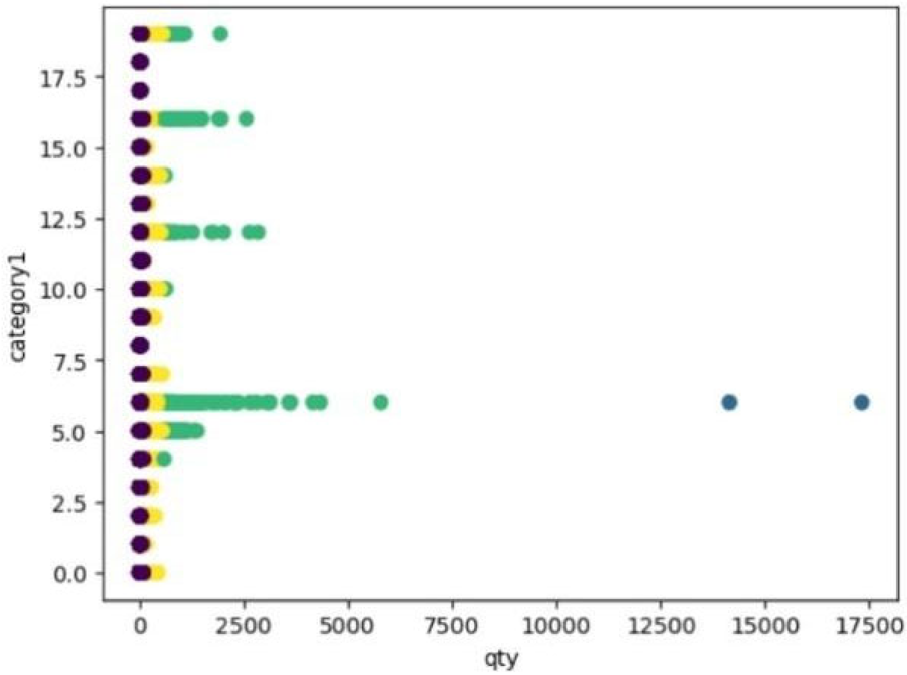

Figure 4. Clustering results(2)

5. Conclusion

We have successfully established the model for predicting the demands and supply from the sales. Time series models can capture the seasonality and trends of historical demand, provide more accurate demand prediction results, and help manage inventory reasonably; The results are relatively interpretable and can help managers understand the reasons for fluctuations in demand; Demand forecasting models based on historical data are usually relatively stable and suitable for long-term supply chain planning. But at the same time, the adaptability to irregular demand is poor: time series models are difficult to cope with certain irregular demand fluctuations, such as increased demand caused by unexpected events; In response to the rapid fluctuations in demand in the short term, time series models may not be flexible enough, resulting in significant errors. For the next step, we can consider introducing more feature engineering, including new dimension attribute information, to improve the demand prediction model. In addition, it is possible to explore how to adaptively update similarity models to adapt to the constantly changing new dimensions.

References

[1] Yifei Zhang.Meituan — China's Leading E-commerce Platform for Local Services Facing the Pandemic[J]. World Scientific Book Chapters, 2023.

[2] Sun Y , Sun Q , Zhu S .Prediction of Shanghai Stock Index Based on Investor Sentiment and CNN-LSTM Model[J]. Journal of Systems Science and Information Technology: English Edition, 2022, 10 (6): 13.

[3] Song J , Zhang L , Xue G ,et al.Predicting Hourly Heating Load in a District Heating System Based on a Hybrid CNN-LSTM Model[J]. Energy and Buildings, 2021, 243(3):110998.

[4] Eshimiakhe D , Lawal K .Application of K-means algorithm to Werner deconvolution solutions for depth and image estimations[J]. Heliyon, 2022, 8(11):e11665.

[5] Wu Y , Zhang Q , Hu Y ,et al.Novel binary logistic regression model based on feature transformation of XGBoost for type 2 Diabetes Mellitus prediction in healthcare systems[J]. Future Generation Computer Systems, 2022, 129:1-12.

[6] Jena R M , Chakraverty S .A NUMERICAL SCHEME BASED ON TWO- AND THREE-STEP NEWTON INTERPOLATION POLYNOMIALS FOR FRACTAL-FRACTIONAL VARIABLE ORDERS CHAOTIC ATTRACTORS[J]. Fractals: An interdisciplinary journal on the complex geometry of nature, 2022(4):30.

[7] Teng X , Zhang X , Luo Z .Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction[J]. Neurocomputing, 2022.

[8] Kilmartin L , Ambikairajah E .A Hybrid Mlp-Rbf Based Speaker Verification System[J]. 2022.

[9] Hamilton W L .Cross-Validation[J].The SAGE Encyclopedia of Research Design, 2022.

[10] Perez R , Kmiecik M , Wilcox S ,et al.ENHANCING THE GEOGRAPHICAL AND TIME RESOLUTION OF NASA SSE TIME SERIES USING MICROSTRUCTURE PATTERNING[J]. 2022.