ISSN 2616-7433 Vol. 5, Issue 18: 104-110, DOI: 10.25236/FSST.2023.051818

Study on Demand Forecasting and Inventory Optimization Using SARIMA and Fuzzy Analysis

Jinyu Liu[1,*], Zhewei Zhang[1], Qiming Song[1], Binyang Luo[1]

[1]Shandong University of Science and Technology, Jinan, China
[*]Corresponding author

Abstract: *With the increasing number of merchants on e-commerce platforms and the continuous enhancement of goods quantity, inventory warehouses in various regions are facing growing warehousing pressure. Therefore, e-commerce platforms need to reduce the warehouse inventory pressure through scientific and effective management means and optimizing the supply chain based on big data application scenarios. This paper analyzes the demand forecasting and inventory optimization issues of e-commerce retail merchants based on the SARIMA model and fuzzy comprehensive evaluation. For the problem, this article first merges the data and organizes the time series data of the same merchant, same warehouse, and same product. Subsequently, the data is processed to remove two outlier values. Considering that the demand for each product is affected by seasonal components, this paper uses the SARIMA model to forecast the time series data and saves the forecast results in the result table. Next, the paper evaluates the forecasting performance of the SARIMA model using fuzzy comprehensive evaluation and calculates the comprehensive evaluation value for the time series data of the same merchant, the same warehouse, and the same product. Based on the calculated comprehensive evaluation value and the 1-wmape index, it concludes that the SARIMA model has good forecasting performance for Problem 1's time series data and the forecasting results are relatively accurate. Finally, this paper classifies the time series with similar comprehensive evaluation values into one category, making the demand characteristics of the same category as similar as possible, and ultimately divides the time series into products with high demand and products with low demand.In conclusion, we evaluated the advantages and disadvantages of the models and methods used in this paper, hoping that future research can further improve these models, enhancing their performance and applicability. Similarly, similar problems can refer to the methods and ideas of this paper for analogous research and analysis.*

Keywords: *SARIMA Model, Fuzzy Comprehensive Evaluation*

1. Introduction

With the increasing number of e-commerce platform merchants and the continuous growth of goods volumes, stock warehouses in various regions are facing greater and greater storage pressure. Therefore, e-commerce platforms need to reduce warehouse inventory pressure through scientific and effective management methods and by optimizing the supply chain based on big data application scenarios. The optimization issues of the supply chain are mainly divided into two aspects: first is demand forecasting, which includes predicting the demand quantity of different products in different warehouses for the region, and on this basis, classifying products with high correlation (similar demand volume)[3]. During the forecasting process, one needs to address issues such as missing data due to short sales periods and unstable sales volumes of certain goods during major promotional events. The second aspect is inventory optimization, with goals to reduce inventory holding costs, meet service levels in a certain region, and lower the turnover days of inventory. Through the aforementioned optimizations, the aim is to reduce storage pressure while satisfying service functions[1-2].

2. Restatement of the Problem

The objective of this study is to enable the e-commerce platform to reduce warehouse inventory pressure through scientific and effective management methods and by optimizing the supply chain based on big data application scenarios. Therefore, based on the demand forecasting and inventory optimization for e-commerce retailers, this paper mainly addresses the following issues: Question:

ISSN 2616-7433 Vol. 5, Issue 18: 104-110, DOI: 10.25236/FSST.2023.051818

Using the data compiled in the past, predict the demand for merchants' goods in various warehouses from May 16, 2023, to May 30, 2023[4], and evaluate the performance of the prediction model. Additionally, discuss how to classify the time series formed by merchants, warehouses, and products so that the characteristics of demand within the same category are most similar.

3. Problem Analysis

The situation requires us to predict the demand for merchants' goods in various warehouses from May 16, 2023, to May 30, 2023, based on the data compiled in the past, and to evaluate the performance of the prediction model used. Additionally, we are to discuss how to classify the time series formed by merchants, warehouses, and products so that the characteristics of demand within the same category are most similar. This paper first preprocesses the data compiled in the past, merging the time series data of the same merchant, the same warehouse, and the same product into one data table to facilitate subsequent analysis and prediction; secondly, this paper uses the SARIMA model to predict the time series data obtained, thus determining the demand for the merchants' goods in various warehouses for the period between May 16, 2023, and May 30, 2023; then, the paper uses fuzzy comprehensive evaluation to assess the performance of the SARIMA model, and according to the results of the fuzzy comprehensive evaluation, classifies time series with similar comprehensive evaluation values into one category, making the characteristics of demand within the same category as similar as possible.The idea of our analysis is shown in Figure 1.

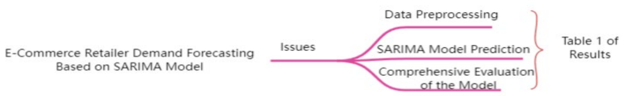


Figure 1: Overall Mind Map of Problem Analysis.

4. Establishment and Solution of the Problem Model

4.1 Data Preprocessing

Firstly, based on the data from Attachment 1, we identified two records as anomalies, which were removed and interpolated by the study. The anomalous data is shown in Table 1**.**

Table 1: Anomalies.

seller_no	product_no	warehouse_no	date	qty
seller_no	product_no	wh_1	2022/12/10	17323
seller_no	product_no	wh_1	2022/12/12	14148

Secondly, this paper uses the pandas library in Python to read and merge data using the merge function, while categorizing it by the same merchant, same warehouse, and same product, and sorting by time.

4.2 Model Establishment

According to the merged data, it is known that the demand for products in various warehouses of merchants constitutes time series data. Considering that the demand for products might be influenced by seasonal elements, this study employs the SARIMA model to forecast the obtained time series data, with the model established as follows:

(a) Establishment of the SARIMA Model

The SARIMA model can be regarded as an extension of the ARIMA model, adding three hyperparameters to define the autoregressive (AR), differencing (I), and moving average (MA) components of seasonality in a time series.

The general form of the SARIMA model can be written as SARIMA(p,d,q)(P,D,Q)s, where p represents the order of the non-seasonal AR term, d is the degree of non-seasonal differencing, q is the order of the non-seasonal MA term, P is the order of the seasonal AR term, D is the degree of seasonal differencing, and Q is the order of the seasonal MA term. Specific model construction is as follows:

(1) AR Model (Autoregressive)

ISSN 2616-7433 Vol. 5, Issue 18: 104-110, DOI: 10.25236/FSST.2023.051818

The autoregressive model only applies to phenomena that are correlated with their own previous values. The mathematical model formula is as follows:

$$y_t = \mu + \sum_{i=1}^{p} r_i y_{t-i} + \epsilon_t \tag{1}$$

The term represents the current value, is the constant term, is the autoregressive coefficient, is the order, and is the error (which is required to follow a normal distribution). The model reflects that there is a linear relationship between the target value at moment (t) and the previous (p) target values, which can be expressed as:

$$y_t \sim r_1 y_{t-1} + r_2 y_{l-2} + .. + r_p y_{t-p} \tag{2}$$

(2) MA Model (Moving Average)

The moving average model focuses on the accumulation of error terms in the autoregressive model. The mathematical model expression is as follows:

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{3}$$

The model reflects a linear relationship between the target value at time t and the previous error values:

$$y_t \sim \theta_1 y_{t-1} + \theta_2 y_{l-2} + .. + \theta_p y_{t-p} \tag{4}$$

(3) ARMA Model (Autoregressive Moving Average)

This model describes the combination of autoregression and moving average, with the specific formula as follows:

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \sum_{i=1}^{p} r_i y_{t-i} \tag{5}$$

Step for Establishing the SARIMA Model:

1) Observe data charts: By plotting the time series chart, observe whether there are trends and seasonality in the data.

2) Stationarity test: Test the time series for stationarity; if not stationary, perform differencing operations until stationary.

3) Fit the model: Determine parameters p, d, q, P, D, Q, and s based on the autocorrelation plot (ACF) and partial autocorrelation plot (PACF) of the stationary data.

4) Model testing: Analyze model residuals to judge the fit of the model.

(b) Fuzzy Comprehensive Evaluation

In Problem 1, there is also a requirement to evaluate the predictive function of the model used and to classify time series so that the demand characteristics within the same category are most similar. This article uses fuzzy comprehensive evaluation to assess the performance of the SARIMA model's predictions, and simultaneously categorizes the time series so that the demand characteristics within the same category are most similar. The specific model establishment steps are as follows:

(1) Determine evaluation indicators: Identify indicators for evaluating the problem and define the evaluation levels or membership functions for each indicator.

(2) Determine membership functions: Define a fuzzy membership function for each indicator's evaluation level, mapping the indicator values to a degree of membership, indicating how much the indicator belongs to a certain evaluation level.

(3) Construct evaluation matrix: Convert the evaluation levels of various indicators into an evaluation matrix, whose elements represent the degree of membership of each indicator at each evaluation level.

ISSN 2616-7433 Vol. 5, Issue 18: 104-110, DOI: 10.25236/FSST.2023.051818

(4) Determine weights: Determine the weights for each evaluation indicator based on the requirements of the problem or expert opinions, to quantify the importance of different indicators.

(5) Fuzzy comprehensive evaluation: Multiply the evaluation matrix by the weights to get a weighted evaluation matrix. Perform fuzzy comprehensive operations on each column of the weighted evaluation matrix (such as maximum, minimum, average, etc.) to obtain a composite evaluation result.

(6) Defuzzification: Process the composite evaluation results for defuzzification, converting fuzzy evaluation results into specific numerical values.

(7) Result analysis and decision-making: Analyze and make decisions based on the defuzzified evaluation results to determine the final evaluation level or make corresponding decisions.

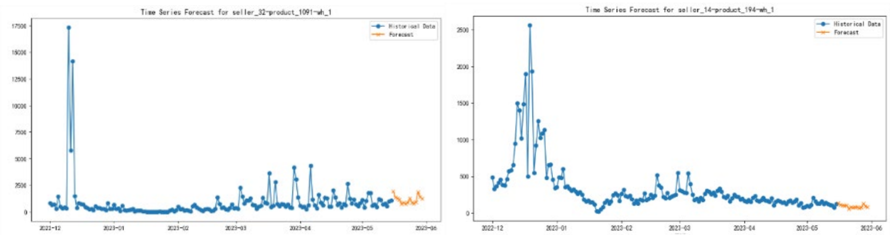
4.3 Model Solution

Table 2: Partial Forecast Results for Problem 1.

seller_32	product_1091	wh_1	2023/5/16	1929.493675
seller_32	product_1091	wh_1	2023/5/17	1370.180479
seller_32	product_1091	wh_1	2023/5/18	1288.065759
seller_32	product_1091	wh_1	2023/5/19	1150.710585
seller_32	product_1091	wh_1	2023/5/20	763.7796232
seller_32	product_1091	wh_1	2023/5/21	880.6070693
seller_32	product_1091	wh_1	2023/5/22	752.6249144
seller_32	product_1091	wh_1	2023/5/23	886.3830395
seller_32	product_1091	wh_1	2023/5/24	1276.203114
seller_32	product_1091	wh_1	2023/5/25	877.4921583
seller_32	product_1091	wh_1	2023/5/26	795.6339927
seller_32	product_1091	wh_1	2023/5/27	903.9445925
seller_32	product_1091	wh_1	2023/5/28	1860.902962
seller_32	product_1091	wh_1	2023/5/29	1395.638388
seller_32	product_1091	wh_1	2023/5/30	1231.779238
seller_23	product_559	wh_1	2023/5/31	618.3445183
seller_23	product_559	wh_1	2023/6/1	619.8631577
seller_23	product_559	wh_1	2023/6/2	596.3419525
seller_23	product_559	wh_1	2023/6/3	576.5611751
seller_23	product_559	wh_1	2023/6/4	599.3234149
seller_23	product_559	wh_1	2023/6/5	599.3064023
seller_23	product_559	wh_1	2023/6/6	558.4307744
seller_23	product_559	wh_1	2023/6/7	560.9210802
seller_23	product_559	wh_1	2023/6/8	552.8933569
seller_23	product_559	wh_1	2023/6/9	582.4061717
seller_23	product_559	wh_1	2023/6/10	579.5376361
seller_23	product_559	wh_1	2023/6/11	603.6331317
seller_23	product_559	wh_1	2023/6/12	624.8351874
seller_23	product_559	wh_1	2023/6/13	626.7541139
seller_23	product_559	wh_1	2023/6/14	605.6644934
seller_14	product_1590	wh_1	2023/6/15	21.71169819
seller_14	product_1590	wh_1	2023/6/16	35.82149322
seller_14	product_1590	wh_1	2023/6/17	55.12398589

By solving the above models, this article obtains the corresponding results, which are stored in the result file, with some results shown in Table 2.

At the same time, this paper merges the forecast results with historical data and visualizes the results, as shown in Figure 2.



ISSN 2616-7433 Vol. 5, Issue 18: 104-110, DOI: 10.25236/FSST.2023.051818

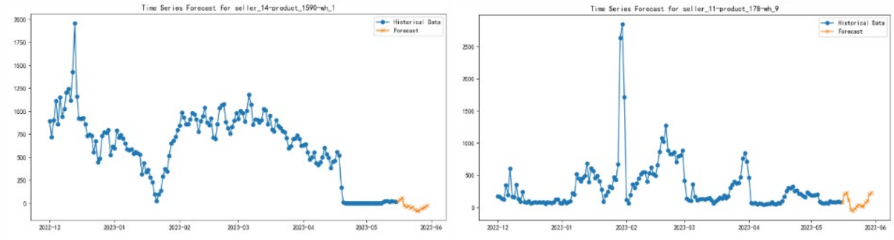


Figure 2: Visualization of Partial Prediction Results.

Subsequently, this paper conducts a fuzzy comprehensive evaluation on the time series data of merchant-warehouse-product and visualizes part of the results as shown in Figure 3.

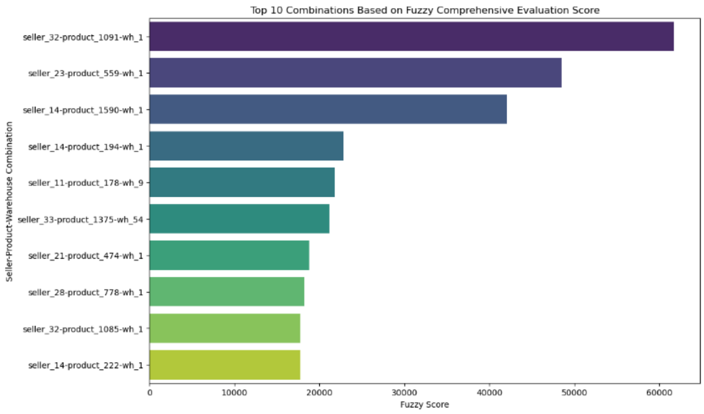


Figure 3: Fuzzy Comprehensive Evaluation Results (Partial).

In the end, time series with similar comprehensive evaluation values are classified into one category to ensure that the demand characteristics within the same category are as similar as possible.The results are shown in Table 3.

Table 3: Classification Results.

Category	qty	seller_category	inventory_category	seller_level	warehouse_category	warehouse_region
1	10.32	8.68	0.91	0.361	0.37	2.65
2	251.1	8.61	1.14	0.46	0.86	2.04

The analysis results show that Category 1 consists of products with smaller demand volumes, primarily concentrated in commodities such as home living and building materials, with warehouses mainly focused in regional storage; Category 2 encompasses products with larger demand volumes, with basic categories also concentrated in commodities such as home living and building materials, and warehouses primarily centralized in central storage facilities.

5. Model Evaluation, Improvement, and Promotion

5.1 Advantages of the Model

(1) SARIMA Model

1) Considers seasonal factors: The SARIMA model can capture seasonal variations in time series data and has good fitting capabilities for data with clear seasonal patterns.

2) Flexibility: The SARIMA model can be adapted to different time series patterns by choosing appropriate parameters, including orders of autoregression (AR), differencing (I), and moving average (MA).

3) Interpretability: The parameters of the SARIMA model have clear statistical meanings, allowing for the interpretation of characteristics of time series data based on the results of parameter estimation.

(2) Fuzzy Comprehensive Evaluation

1) Capable of handling issues of uncertainty and fuzziness, suitable for real applications with incomplete information and vague concepts.

ISSN 2616-7433 Vol. 5, Issue 18: 104-110, DOI: 10.25236/FSST.2023.051818

2) Allows for a comprehensive integration of multiple evaluation indicators to obtain more holistic and accurate evaluation results.

3) The algorithm is relatively simple, easy to implement and apply.

4) Parameters of the fuzzy comprehensive evaluation can be flexibly adjusted to meet different evaluation needs.

5.2 Disadvantages of the Model

(1) SARIMA Model

1) High data requirements: The SARIMA model demands high-quality data that satisfy conditions of stationarity and seasonality; otherwise, the fit of the model may be poor.

2) Difficulty in parameter selection: Choosing parameters for the SARIMA model requires observation of autocorrelation (ACF) and partial autocorrelation (PACF) plots, which can be challenging for complex time series patterns.

(2) Fuzzy Comprehensive Evaluation

1) The method of fuzzy comprehensive evaluation involves complex modeling and parameter setting, requiring professional knowledge and experience.

2) When dealing with large-scale data and complex issues, the computational complexity is high, necessitating long computation times.

5.3 Improvements and Extension of the Model

(1) SARIMA Model

1) The SARIMA model could be further extended to the SARIMAX model, including external variables as additional explanatory variables to enhance prediction capabilities.

2) Other more complex seasonal models, such as seasonal decomposition or seasonal exponential smoothing models, could be tried to improve modeling and prediction accuracy for seasonal variations.

3) Machine learning methods, like neural networks or deep learning models, could be combined to improve modeling and forecasting abilities for time series.

4) Methods such as cross-validation could be used to evaluate model performance, facilitating model selection and parameter tuning.

(2) Fuzzy Comprehensive Evaluation

1) Further research and improvements could be made on the modeling and parameter setting of the fuzzy comprehensive evaluation method to increase its applicability and accuracy.

2) It could be combined with other evaluation methods, such as Analytic Hierarchy Process (AHP) or Fuzzy AHP, for a more comprehensive evaluation to improve the credibility and accuracy of the results.

3) Machine learning and artificial intelligence technologies could be applied to optimize the computational efficiency and accuracy of the fuzzy comprehensive evaluation method through big data analysis and automated learning.

4) The fuzzy comprehensive evaluation method could be applied to a broader range of fields, such as finance, healthcare, environment, etc., to expand its application scope and impact.

6. Conclusions

This article aims to address the issue of reducing warehouse inventory pressure in e-commerce platforms by applying big data to supply chain scenarios. By employing scientific and effective management methods along with data analysis-based forecasting models, this paper will predict merchants' product demand in different warehouses and assess the performance of the forecasting models. Additionally, the paper explores how to categorize the time series formed by merchants, warehouses, and products so that the demand characteristics within the same category are as similar as

ISSN 2616-7433 Vol. 5, Issue 18: 104-110, DOI: 10.25236/FSST.2023.051818

possible.

In addressing the problem, we will use the data compiled in the past to forecast the merchants' product demand for the period from May 16, 2023, to May 30, 2023, and evaluate the performance of the forecasting models.

References

[1] Huang Yi, Yang Yongsheng, Zhang Qingguang. Application of the SARIMA Model in Monthly Average Temperature Time Series [C]// Strengthening Scientific and Technological Foundations to Promote Meteorological Modernization—The 29th Annual Meeting of the Chinese Meteorological Society. 2012.

[2] Han Zhonggeng. Mathematical Modeling Methods and Applications [M]. Higher Education Press, 2017.

[3] Zhang Li, Niu Huifang. Forecast Analysis of Consumer Price Index Based on SARIMA Model [J]. Journal of Applied Statistics and Management, 2013, 32(1):6. DOI:CNKI:SUN:SLTJ.0.2013-01-004.

[4] Yang Jinfang, Zhai Yongjie, Wang Dongfeng, et al. Time Series Prediction Based on Support Vector Regression [J]. Proceedings of the CSEE, 2005, 25(17):5. DOI:10.3321/j.issn: 0258-8013. 2005. 17.022.