Research on Express Demand Forecasting Based on the Entropy Weight-TOPSIS and ARIMA Models

Hanxin Chen*

Faculty of Engineering, Environment and Computing, Coventry University, Priory Street, Coventry, United Kingdom

chenhanxin2000@163.com

*Corresponding author

Keywords: Express demand forecasting, Entropy Weight-TOPSIS model, ARIMA model, Logistics network, City importance evaluation, Time-series analysis

Abstract: With the rise of e-commerce platforms and the popularity of online shopping, the express delivery industry is experiencing rapid demand growth. Accurate demand forecasting has become crucial for optimizing logistics networks, reducing costs, and enhancing service quality. This paper proposes a comprehensive framework for express demand forecasting based on the Entropy Weight-TOPSIS and ARIMA models. First, the Entropy Weight-TOPSIS model is used to evaluate the importance of cities in the express logistics network. Six indicators, including the number of supply cities, the number of receiving cities, average shipment volume, average receipt volume, shipment change rate, and receipt change rate, are selected to rank 24 cities comprehensively. The results show that cities L, G, V, W, and B have the highest importance in the express logistics network. Subsequently, the ARIMA model is employed for time-series analysis and forecasting of express demand. Through time-series modeling of the total express volume, combined with ADF tests and autocorrelation analysis, an ARIMA(0,1,2) model is established to predict express transportation volumes for the next two days. The model exhibits good forecasting performance with a fit of R^2 = 0.837. The study demonstrates that the integrated approach combining the Entropy Weight-TOPSIS and ARIMA models can effectively evaluate the importance of cities in the express logistics network and provide accurate references for express demand forecasting, offering strong support for logistics planning and resource allocation by express companies.

1. Introduction

In recent years, the rapid development of e-commerce has led to a significant increase in the demand for express delivery services [1]. Online shopping has become a dominant trend in modern consumer behavior, driving the express delivery industry to expand at an unprecedented rate. As a result, accurate demand forecasting has become essential for express delivery companies to optimize their logistics operations, allocate resources efficiently, and enhance service quality. Traditional forecasting methods, such as historical averages and simple statistical models, are no longer sufficient to handle the complexity and volatility of modern express delivery demand. Therefore, there is a need for more advanced and sophisticated forecasting techniques that can integrate multiple data sources and account for various influencing factors, such as seasonal changes, promotional activities, and real-time market dynamics. This paper aims to address this challenge by proposing a novel forecasting framework that combines the Entropy Weight-TOPSIS and ARIMA models.

2. Literature Review

With the rise of e-commerce platforms and the popularity of online shopping, the express delivery industry is facing unprecedented demand growth. In order to respond effectively to this change, demand forecasting methods are also evolving, from traditional statistical methods to more sophisticated modern techniques [2].

2.1. Evolution of Traditional and Modern Forecasting Methods

Abolghasemi et al and Petropoulos et al research found that early demand forecasting mainly relied on simple statistical methods [3, 4], such as historical average, moving average or exponential smoothing, which could provide a certain degree of forecasting accuracy when the data was relatively stable and the model was relatively simple. However, as the complexity of the market increases, the predictive accuracy of these traditional methods is limited due to the lack of ability to handle complex data patterns [3].

With the rapid development of information technology, especially the application of big data and artificial intelligence technology, express demand forecasting methods have also ushered in new developments and opportunities. For example, ARIMA models for time series analysis, multi-layer perceptron neural networks, and linear programming models are modern forecasting techniques that can handle more complex data relationships and capture unpredictable market changes to improve the accuracy and efficiency of forecasts [5].

2.2. New Rends Predicted by the Express Delivery Industry

In the express delivery industry, especially in the rapid development of e-commerce today, the demand forecast for express delivery is more urgent and complex. On the one hand, predictive models need to have the ability to process massive data and capture complex trends. On the other hand, it is also necessary to take into account the impact of seasonal changes, promotional activities and other factors on express demand.

In order to meet these challenges, some advanced forecasting techniques and methods have been gradually applied. For example, by combining the advantages of multiple forecasting models, model fusion technology is used to improve the overall forecasting performance. Dynamic demand forecasting with real-time data analytics and cloud computing to more flexibly adjust logistics strategies and resource allocation [6]. In addition, the introduction of deep learning and artificial intelligence technologies has brought greater accuracy and insight to delivery demand forecasting.

In summary, with the progress of technology, express demand forecasting is gradually developing towards a more intelligent and refined direction. These advances not only improve the accuracy of the forecast, but also provide strong technical support for express logistics optimization, cost control and service quality improvement [7].

3. Evaluating the Importance of Cities in Express Logistics Network

The importance of the cities of each site is comprehensively sorted. First, data processing and analysis are required, and the relevant important degree indicators are selected according to the results obtained by the statistical analysis, to explore the similarities and differences between the cities. Subsequently, the relevant evaluation method was used to analyze them weighted, and a comprehensive evaluation model was established to comprehensively rank the cities of each site.

3.1. Selection and Calculation of the Evaluation Index

The evaluation index should depict the characteristics or attributes of the evaluated object. When selecting indicators, factors such as their relevance, mutability and practicality should be considered. We select six indicators to comprehensively reflect the importance of each city in express transportation from different angles.

First, the number of supply cities and the number of receiving cities reflect the connectivity of a city in the express network, that is, the degree of the city depends on the express transportation demand and supply of other cities. These two indicators can help identify cities with high connectivity in the network, and then determine the core sites of express transportation. This is an important reference factor for express delivery companies to develop logistics routes and the layout of warehouse sites.

Secondly, the average shipment and the average receipt reflect the size of the express delivery volume in a city. These two indicators can help understand the transportation volume of each city, provide accurate business data analysis for express delivery companies, so as to carry out targeted logistics planning and resource allocation, and further optimize logistics efficiency and service quality.

Finally, the average daily rate of change and the average daily rate of delivery reflect the changing trend of urban express delivery volume. These two indicators can help to understand the change law of express transport volume between cities, and timely adjust logistics resources and operation strategies to adapt to the changes. At the same time, these two indicators can also help to predict the future transport demand, and further optimize the logistics planning and resource allocation.

Therefore, selecting these six indicators can improve the operation efficiency and service quality of urban logistics, and provide an important reference for them to formulate reasonable logistics routes and the layout of warehouse sites.

3.2. Data Processing and Visualization

Calculating the number of supply cities and receiving cities requires traversing each city and counting the number of different cities in its transport/receiving cities list. Specifically, we use the dictionary data structure in Python. To achieve this statistical function, take the name of the city as the key of the dictionary, the list of delivery / receiving cities as the value of the dictionary, and use the set data structure to weight, and finally count the number of different cities is the number of supply / acceptance cities of the city.

For the average value of shipments and the average amount of deliveries, first adds up the number of express deliveries in each city in Dataset to get the total shipment and the total receiving quantity, and then divided by the number of deliveries and deliveries in the city to get the average value.

The average daily average rate of change of shipments and the daily average rate of delivery can be calculated from the daily shipments and deliveries of each city in Annex I. Specifically, we group the delivery date in Annex 1 by day. For each city and each day, it accumulates the shipment / receiving volume of the city on that day, and statistics the shipment / receiving times of the city in that day. Then, for each city, it obtains the fitting coefficient, which is the average change rate of shipment / receiving volume of the city.

Through the above solution, the specific data of the six indicators are compiled, as shown in Table 1.

Table 1 The significance of urban logistics.

city	Supply city quantity	Receive city quantity	The shipment volume mean value	Receipt quantity mean value	Shipments change rate	Receiving goods change rate
A	2	4	82.089	63.213	0.011166	0.04908
B	1	1	185.94	236.727	0.102053	0.205003
C	4	2	33.814	37.687	-0.00743	-0.04777
…	…	…	…	…	…	…
W	3	3	146.635	161.453	0.164266	0.206398
X	4	5	99.154	91.534	0.11336	0.119373
Y	3	2	90.806	116.791	0.104243	0.154054

In this paper, radar map and line chart are drawn for visual analysis of the daily average change rate of supplied and accepted cities and received quantity, as shown in Figure 1.

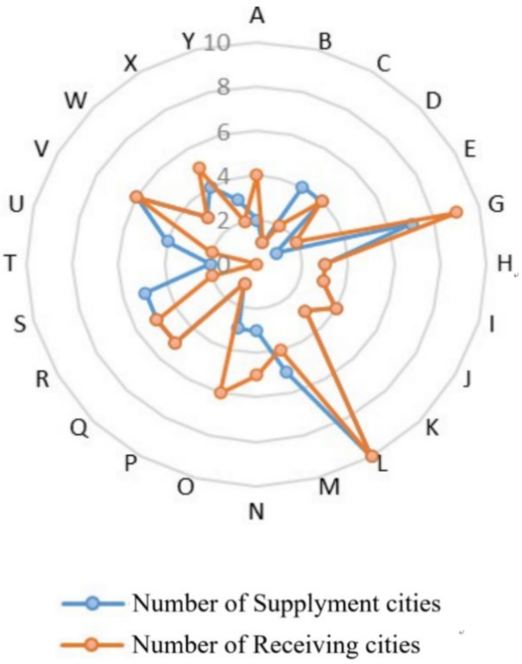

Figure 1 Radar diagram of station supply city and receiving city quantity.

According to the radar chart, the number of supply cities and accept cities is very similar. Among them, L City and G City stand out, as they have the highest numbers for both supply and accept cities, indicating that they are important transport hubs. In contrast, Q, K, J, and I cities have a supply count of 0, while T city also shows no supply activity. Additionally, B, E, and Q cities have relatively low accept city numbers. Notably, L City has the highest supply and accept numbers, reaching 10.

We also make a comparison of the change rate of delivered quantity and received quantity in Figure 2.

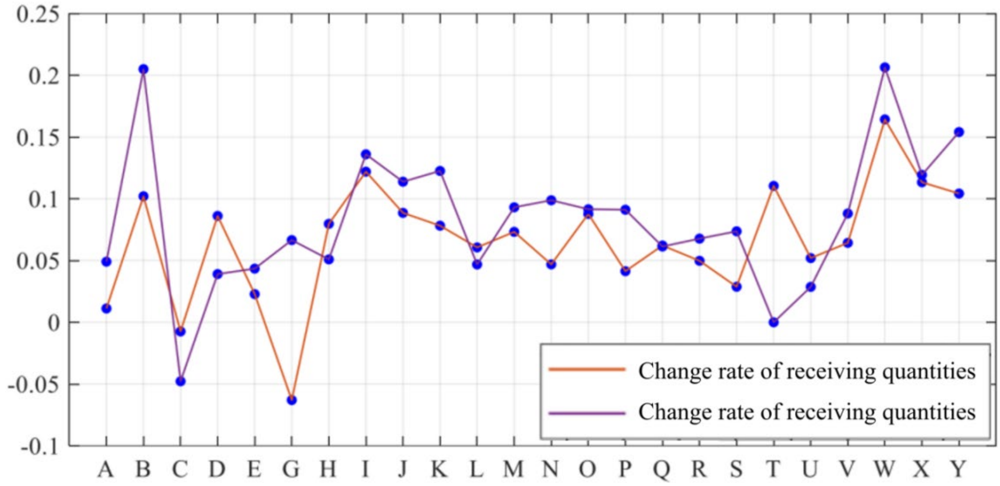

Figure 2 Average rate of change of the received and shipped quantity.

As can be seen from the chart above, the growth rate of C and G cities is negative, while the growth rate of other cities is positive or unchanged. As can be seen from the chart above, the growth rate of B cities and W cities is relatively high, with a growth rate of more than 20%. It can also be seen from the curve in the figure that except for G and T cities, the delivery growth rate and the delivery growth rate change in the opposite direction, while the receiving growth rate and the delivery growth rate of other cities all change in the same direction.

3.3. Entropy Weight-TOPSIS Model Building

Entropy weight method is a common index weight determination method, which can help to determine the influence degree of each index on the evaluation results, so as to ensure that the evaluation results are more objective and accurate. However, TOPSIS method is a commonly used multi-index comprehensive evaluation method. It can find the scheme closest to the optimal solution, to conduct comprehensive ranking. The model was evaluated using the entropy weight method-TOPSIS [8]. Wecan consider the influence of supply city quantity, city quantity, delivery average, delivery, delivery and daily average rate of change, the importance of each site city objective accurate evaluation and sorting, for the express company warehouse site layout, storage cost saving and transportation route planning to provide strong support and reference, has important practical significance. The specific process is as shown in Figure 3.

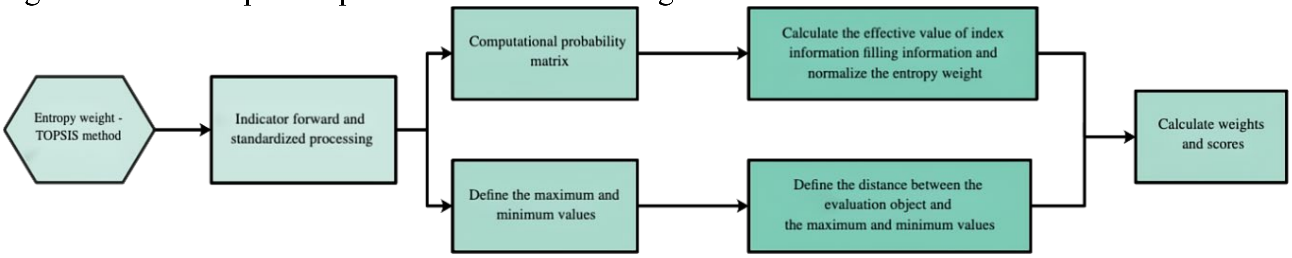

Figure 3 Flow chart of entropy weight-TOPSIS method.

Initially, the raw data matrix is standardized to eliminate the influence of different units and scales. This standardization ensures that all indicators are comparable. Subsequently, the entropy value for each indicator is calculated to measure the degree of disorder or information content within the data. Based on these entropy values, the entropy weight coefficients are derived, reflecting the relative importance of each indicator. These weights are then used to compute a weighted normalized matrix by multiplying the standardized data with the weight vector. The next step involves identifying the optimal and worst solutions within the weighted normalized matrix. Distances from each alternative to both the optimal and worst solutions are calculated. Finally, a composite score is determined for each alternative based on these distances, with higher scores indicating better performance. The alternatives are then ranked according to their composite scores, allowing for a clear prioritization of the most favorable options.

3.4. Solution of the Entropy Weight-TOPSIS Model

According to the above entropy weight-TOPSIS evaluation model, we process and analys the provided data, and obtains the comprehensive evaluation index of each site city. By ranking these indices, we can comprehensively rank the importance of each site city and get the top 5 site cities in the importance ranking.

As shown in Table 2, the utility value of the accept city number is influenced by two primary variables: the supply city quantity and the delivery quantity rate, the latter of which has the smallest weight. Among the factors analyzed, the supply and accept quantities carry the highest weight, followed by the average values for both supply and accept cities. Finally, the delivery quantity change rate and the accept quantity rate have the lowest weights.

Table 2 Entropy weight method.

metric	Information entropy value e	Information utility value d	weight (%)	metric	Information entropy value e	Information utility value d	weight (%)
Average shipment value	0.937	0.063	18.709	received cities value	0.913	0.087	26.143
Average received goods value	0.962	0.038	11.415	Chang rate of shipment quantity	0.973	0.027	7.946
Supply cities value	0.915	0.085	25.508	Chang rate of received quantity	0.966	0.034	10.279

The positive ideal solutions in Table 3 above all are close to 1 and have the best of each alternative;

the average shipment and receiving average are the smallest and the worst of the alternatives.

Table 3 TOPSIS positive and negative ideal solutions.

metric	Positive ideal solution	Negative ideal solution	metric	Positive ideal solution	Negative ideal solution
Average shipment volume	1	4.70E-07	Number of receiving goods	0.999989	1.11E-05
Average goods received value	1	4.20E-07	Delivery rate	0.99956	0.00044
Number of shipments	0.999989	1.11E-05	Receiving rate	0.999607	0.000393

The comprehensive score of the 24 cities was calculated by the weight of each factor, and it was ranked according to the comprehensive score, thus obtaining the following figure (see Appendix B for detailed data). As can be seen from Figure 4, the top five cities are L, G, V, W and B.

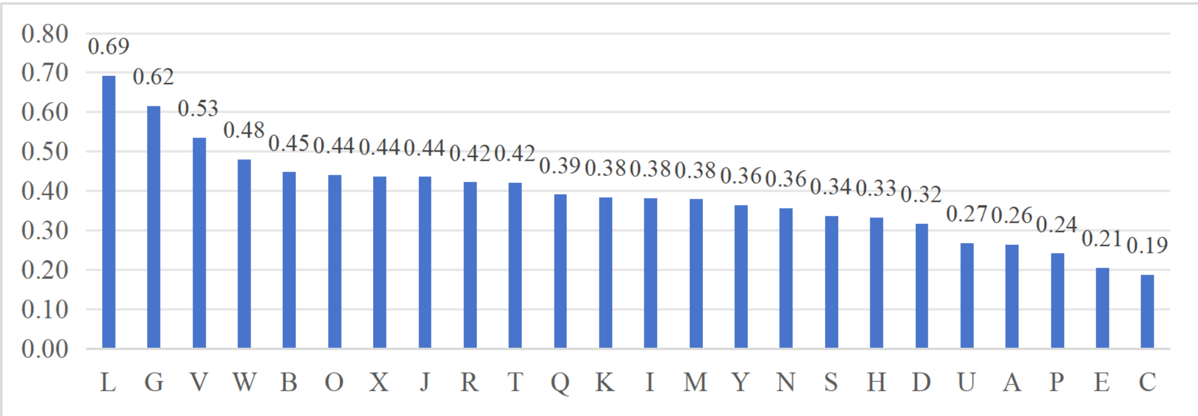

Figure 4 Solution for evaluating the importance of cities.

4. Evaluating the Importance of Cities in Express Logistics Network

Time series analysis method can use historical data and trends to predict future values, so it can provide more accurate prediction results. At the same time, data can be processed in real time and predicted, so that enterprises can make timely decisions. Through the visual display of the prediction results, enterprises can better conduct data analysis and make decisions. We use time series analysis to predict express delivery quantity. First, the total number of express deliveries per day in the data was summarized for time-series model prediction. Then, the number of express deliveries between various sites was classified and summarized, followed by time series model prediction. Due to the limited space and to avoid repeating, we only describe the time series model with the total number of "delivery-receiving" city express transportation samples calculate. See the attachment for the relevant data of the time series prediction model of express delivery quantity between the cities of "delivery-receiving" station in the city.

4.1. Establish the Time Series

A time series was first established based on the total number of deliveries during this period. Subsequently, a stability analysis was conducted using the Augmented Dickey-Fuller (ADF) test, which is a common method for determining whether a time series is stationary or contains a unit root. The ADF test was performed using SPSS software, and the results were visualized through relevant time series plots. The ADF test data provided insights into the stability characteristics of the delivery time series, helping to identify any trends or patterns that may affect further analysis. Table 4 is the relevant data of the ADF test.

Table 4 The ADF Test Form.

variable	Differential order	t	P	AIC	critical value		
					1%	5%	10%
Express delivery quantity(PCS)	0	-3.564	0.006***	4512.836	-3.455	-2.873	-2.573
	1	-7.703	0.000***	4504.376	-3.456	-2.873	-2.573
	2	-7.427	0.000***	4529.489	-3.456	-2.873	-2.573

Note: * * *, * * and * represent the significance levels of 1%, 5% and 10%, respectively.

As can be seen from the above table, when the difference is order 0, the P-value is less than 0.05, indicating that the rejection of the null hypothesis and the sequence is a stationarity test (Tian & Wu, 2022). The same t-value also shows that the sequence is a stability test; when the difference is in order 1, the P value is less than 0.05, indicating that the original hypothesis is rejected, the sequence is a stability test, and the similar t value is a stability test; when the difference is in order 2, the P value is less than 0.05, indicating that the original hypothesis is rejected, the sequence is a stability test, and the same t-value also shows that the sequence is a stability test. At the same time, make a time sequence diagram of the time series, as shown in Figure 5.

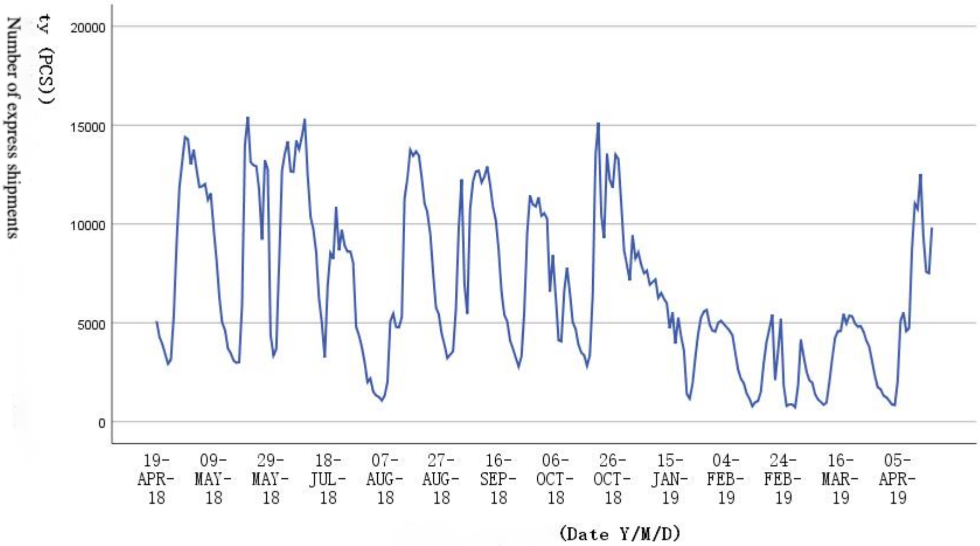

Figure 5 The timing diagram.

As can be seen from the change of the above figure over time, this time series figure is relatively smooth, and a relevant model can be built.

4.2. Establish the ARIMA model

An autocorrelation plot was made using MATLAB, as shown in Figure 6. From the correlation graph, we know that it tends to 0 rapidly after greater than a constant K, and three samples exceed the given confidence interval, and the correlation coefficient of this graph is sored.

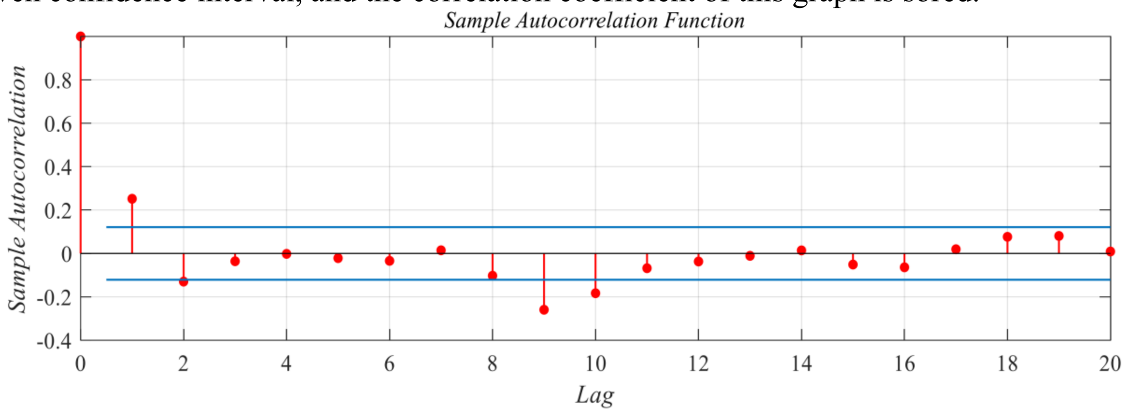

Figure 6 Autocorrelation plots.

The partial autocorrelation diagram is also made as shown in Figure 7. Therefore, the ACF decreases sharply, and the partial autocorrelation coefficient of this graph is dragging.

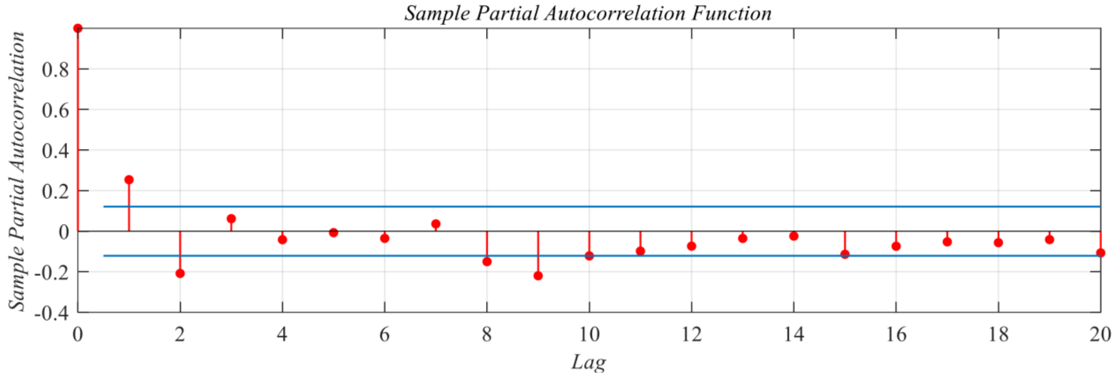

Figure 7 Partial autocorrelation plots.

From the autocorrelation diagram and partial autocorrelation diagram, ACF is cut at lag 2, and PACF is trailed, so the time series ARIMA (0,1,2) model was selected to establish the formula as follows:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{1}$$

From the analysis of the Q statistic results, it is possible that Q6 does not show significance at the level.as The assumption that the residual is white noise sequence cannot be rejected. The goodness of fit of the simultaneous model R^2=0.837, the model performance is excellent, and the model basically meets the requirements. The ARIMA model parameters can be made as shown in Table 5.

Table 5 Table of model parameters.

	coefficient	standard error	t	P>|t|	0.025	0.975
constant	17.423	118.621	0.147	0.883	-215.07	249.916
L1 Express delivery quantity (pieces)	0.318	0.061	5.188	0	0.198	0.439
L2 Express delivery quantity (pieces)	-0.126	0.063	-2.012	0.044	-0.249	-0.003

The P-value shows that the significance of the two parameters is less than 0.05, so the formula of the model is:

$$Y(t) = 17.423 + 0.318 * \varepsilon(t-1) - 0.126 * \varepsilon(t-2) \tag{2}$$

Make a time series diagram with SPSS, as shown in the Figure 8.

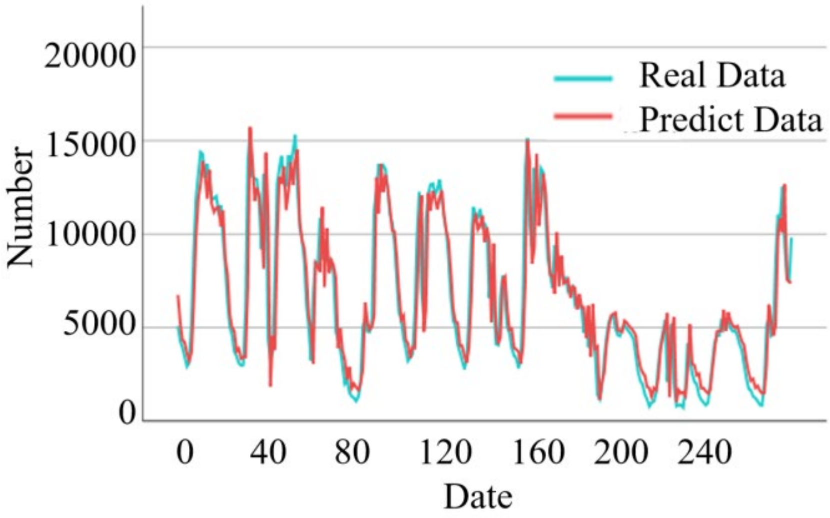

Figure 8 Time series Figure.

According to the prediction formula of the time series, we can predict the number of express transportations between the cities of each "ship-receiving" station on April 18,2019 and April 19,2019, and the total number of express transportation between all the "ship-receiving" station cities on the same day, as shown in Table 6.

Table 6 Results of express demand forecasting.

	coefficient	standard error	t	P>|t|	0.025	0.975
constant	17.423	118.621	0.147	0.883	-215.07	249.916
L1 Express delivery quantity (pieces)	0.318	0.061	5.188	0	0.198	0.439
L2 Express delivery quantity (pieces)	-0.126	0.063	-2.012	0.044	-0.249	-0.003

5. Conclusion

In conclusion, the entropy weight-TOPSIS method has proven to be an effective tool for evaluating and ranking alternatives based on multiple criteria. It successfully integrates standardized data treatment, entropy-based weight calculation, and comprehensive scoring to provide a clear and objective ranking of options. However, there is room for improvement. Future work could focus on incorporating additional indicators to enhance the model's comprehensiveness and adapting it to dynamic data environments to ensure more robust and timely decision-making.

References

[1] Tekin M, İnce H, Etlioğlu M, et al. A study about affecting factors of development of e-commerce[C]//Proceedings of the International Symposium for Production Research, 2018. Springer International Publishing, 2019: 625-642.

[2] Syntetos A A, Babai Z, Boylan J E, et al. Supply chain forecasting: Theory, practice, their gap and the future[J]. European Journal of Operational Research, 2016, 252(1): 1-26.

[3] Abolghasemi M, Beh E, Tarr G, et al. Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion[J]. Computers & Industrial Engineering, 2020, 142: 106380.

[4] Petropoulos F, Makridakis S, Assimakopoulos V, et al. ‘Horses for Courses’ in demand forecasting[J]. European Journal of Operational Research, 2014, 237(1): 152-163.

[5] Lim B, Zohren S. Time-series forecasting with deep learning: a survey[J]. Philosophical Transactions of the Royal Society A, 2021, 379(2194): 20200209.

[6] Nowicka K. Smart city logistics on cloud computing model[J]. Procedia-Social and Behavioral Sciences, 2014, 151: 266-281.

[7] Ittmann H W. The impact of big data and business analytics on supply chain management[J]. Journal of Transport and Supply Chain Management, 2015, 9(1): 1-9.

[8] Zhiyong Y, Yuxi Y, Yu Z. Application of BiLSTM-SA-TCN lime series model in stock price prediction [J]. Journal of Nanjing University of Information Science and Technology, 2023, 15(6): 643-651.