ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0081

A Review of Rail Freight Forecasting Studies

Wei Wang

School of Business Administration, Henan Polytechnic University, Jiaozuo 454000, Henan, China

Abstract

To enhance the efficiency of railway freight organization and improve the foresight of railway logistics planning, this paper reviews the mainstream railway freight volume forecasting methods at the present stage. Firstly, based on analyzing the current situation of the research on the influencing factors of railway freight volume, the methods for analyzing the influencing factors of railway freight volume are summarized; Secondly, the methods applied in the field of railway freight volume forecasting are divided into two categories: traditional forecasting methods based on statistics and intelligent forecasting methods based on machine learning and analyzed for application, focusing on the principles and application scenarios of different methods, and the common forecasting under different methods The objective review of common forecasting models under different methods is also presented. Finally, potential research hotspots and difficulties in the field of railway freight forecasting are identified based on the current research status and development characteristics, to provide references for railway freight organization and operation.

Keywords

Freight Forecasting; Railway Freight; Forecasting Model.

1. Introduction

Rail freight forecasting refers to the process of forecasting the development trend of rail freight volumes within the planning horizon based on the historical data of rail freight volumes and the relevant data on rail freight development[1]. In recent years, the rapid development of international trade, along with the development trend of economic trade, supply chain globalization and "One Belt, One Road", "public to rail" and other national policies to boost, railway transport in the creation of green logistics system, to achieve China's economic transformation The mainstream advantages of railway transport are gradually revealed. Therefore, scientific forecasting of railway freight volume based on statistical data will have an important role in promoting the adjustment of China's overall transport structure, the improvement of the overall management level of railway freight, and the optimization of social resource allocation.

The railway transport system itself has extremely complex and dynamic characteristics and is an ecosystem that works through a combination of factors. This uncertainty as well as ambiguity brings new challenges to the accuracy of scientific forecasting. In general, forecasting using traditional models requires accurate raw data series, and the process is mathematically modeled through deductive reasoning. In the current environment, however, the use of traditional models is increasingly limited and there is an urgent need to find more stable, accurate and robust models to advance the forecasting of rail freight volumes to reduce logistics costs and provide a more solid database for the organization and operation of the rail freight system.

This paper analyses the results of several domestic and international research studies, including not only the different factors affecting rail freight volumes, but also the current state of research on rail freight forecasting, and analyses the different methods used in the process of forecasting rail freight volumes, summarises the latest forecasting methods and combined models and provides an outlook on future trends in rail freight forecasting.

2. Methodology for Analyzing Factors Influencing Rail Freight Volumes

In the process of analyzing the factors that affect rail freight volumes, two main methods are used: quantitative analysis and qualitative analysis.

2.1 Quantitative Analysis Method

The quantitative analysis method mainly uses positivism as the theoretical foundation for the act of judging things through their internal and their inherent logical causal links. The use of quantitative analysis to determine the factors affecting rail freight volumes is generally characterized by VAR models[2], linear regression[3], grey correlation[4] and other analytical methods.

VAR, also known as the value at risk model, is generally used in the financial industry for risk assessment and is based on the various possibilities of predicting the existence of data under the action of multiple variable parameters. According to Granger's theory, the correlation between the three, freight volume, passenger volume and GDP, was found to have no direct relationship between passenger volume and GDP after data analysis using the SVR approach, while there is a lasting equilibrium relationship between freight volume and GDP levels. Without taking full account of the time series characteristics, the use of traditional statistics would infer a close relationship between passenger traffic and GDP. The VAR model adds the time characteristics of the series to the variable factors, achieving the ability to reflect the overall variables dynamically, although based on the limitations of the unstable nature of this model algorithm, its predictions are subject to some bias as environmental differences change. Linear regression is a statistical analysis method that uses the regression method of mathematical statistics to determine the interdependence between two or more variables and is often used in areas such as finance, technology, and logistics where forecasting is required. The limitations of multiple linear regression analysis include the lack of consideration of interaction effects and non-linear causality, but it has the advantage of determining the factors influencing freight volumes based on an accurate measure of the degree of correlation between multiple variables and the degree of regression fit. Grey correlation analysis uses grey systems theory to deal with limited information and to derive correlations between various random variables based on the serial curves presented by the data. In the early stages of forecasting freight volumes, the use of grey correlation analysis can greatly improve the accuracy of the overall model.

By analyzing the application of the above three methods, the factors affecting the results of rail freight forecasting can be summarised as follows.

(1) The use of VAR models, with their hidden assumptions, is less appropriate for railway freight systems where there are multiple possibilities and obvious trends of change; linear regression analysis is limited by a relatively simple algorithmic model, which prevents the further in-depth study of more complex systems; grey correlation analysis is suitable for situations where the initial series is crude and incomplete.

(2) In addition to this, we must also take into account the changing times and the changing environment of policy adjustment. According to the existing research results, railway freight traffic is mainly influenced by a series of factors such as policy changes, information tools, technology levels, infrastructure road networks, national economic levels and international trade, etc. To improve the accuracy of forecasting, we need to take into account the current context and environmental policies.

2.2 Qualitative Analysis Method

Qualitative analysis is a method in which the forecaster makes a subjective analysis of the nature or trend of something based on his knowledge of the present or development trend based on the forecaster's knowledge of the method.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0081

Qualitative analysis uses a non-quantitative approach to determine the nature of things It is important to consider the level of expertise of the forecaster, the availability of historical information, and the stability of the It requires a high level of expertise, the availability of historical information and the stability of the environment. As a result, cannot fully reflect the complexity of the rail freight system and is limited to subjective The complexity of the rail freight system is not fully reflected and is limited to a subjective judgment of whether it is an influencing factor on the volume of rail freight, rather than a scientific analysis of its correlation. It is limited to a subjective judgment of whether or not it is an influencing factor on rail freight volumes, rather than a scientific analysis of the relationships.

3. An Overview of Rail Freight Forecasting Methods

Forecasting rail freight volumes requires sufficient historical data and the creation of scientific models to achieve accurate forecasts. Currently, qualitative, quantitative and combined forecasting methods are commonly used in railway freight forecasting research with outstanding results.

3.1 Qualitative Forecasts Method

As a whole, qualitative forecasting is a comprehensive and discursive process that requires the use of comparative argumentation to make inferences that are not constrained by schematic procedures and are characterized by a high degree of tension and subjectivity. Qualitative forecasting is generally carried out using pooling opinions, brainstorming and comparative analogies.

In contrast to the traditional approach, qualitative forecasting is becoming an important method throughout the forecasting process due to the influential role of the nature of fundamental thinking. In analyzing the factors that affect rail freight volumes, it is necessary to use the brainstorming of qualitative analysis; in the process of quantitative research of data using mechanized procedures and data control procedures, the use of qualitative forecasting methods prevents quantitative analysis from being reduced to pure data calculation. This has led to the gradual development of forecasting research based on the foundational tools of qualitative research, the creation of basic data forecasting models, and the full combination of quantitative and qualitative strengths for scientific forecasting of rail freight volumes.

3.2 Quantitative Forecasts Method

Quantitative forecasting refers to the analysis and study of the patterns of change in data series through the creation of statistical and mathematical models for existing data, and the prediction of trends in rail freight volumes based on the study of the patterns. The use of quantitative forecasting methods is common in the two categories of traditional forecasting based on statistics and modern forecasting based on machine intelligence.

3.2.1 Traditional Forecasting based on Statistics

Traditional forecasting based on statistics means that existing mathematical theories are used to predict in advance and deductive reasoning is applied to the entire forecasting process, generally using two methods of time series and causal linkage for extrapolation forecasting.

(1) Time series forecasting

Based on the characteristics of one-dimensionality of time, time series analysis is used to study the pattern of changes before and after, to predict the future development trend. Common methods include moving average, exponential smoothing, seasonal index forecasting, trend extrapolation and the Box-Jenkins method.

For example, the moving average and exponential smoothing methods have been applied to rail freight forecasting, and their findings show that the overly homogeneous traditional forecasting models not only suffer from low accuracy but also show weaknesses in extrapolation. The use of Holt-Winter to forecast rail freight volumes highlights the advantages of this model in predicting trends in freight volumes over shorter periods for a single category of goods[5]. A time series with

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0081

small fluctuating trends and insignificant changes in the data series is known as a smooth series. In contrast to the use of ARMA for the analysis of smooth series, non-smooth series can be analyzed using a differential integrated moving average autoregressive model. The Holt-Winter model is very efficient in forecasting both trend and seasonal time series, with the advantage of moving average and full period averaging, while the ARIMA model relies on endogenous variables and does not require exogenous driving variables, achieving higher accuracy in forecasting rail freight volumes in the short and medium-term[6].

In summary, the use of time series forecasting does not require a high level of model creation and computing power, but the premise is that the analysis must be based on the preservation of more complete historical data, in addition to the extrapolation of current environmental trends, once other dynamic or sudden influences are ignored, the model will lead to forecasting deficiencies, which means that the model will also be less effective in forecasting freight volume trends in the medium to long term. This means that the model will be less effective in predicting freight volume trends in the medium to long term.

(2) Causal analysis forecasting

Causal analysis forecasting is based on positivism, which is the analysis of a series of causal relationships between things to deduce the future trend of the development of things. The methods used in railway freight forecasting mainly include linear and non-linear regression forecasting. Causal analysis forecasting has also been widely used in various railway freight volume forecasting studies. The results of a study carried out to forecast and analyze freight demand over the planning horizon based on six months of freight volume data for the port of Mumbai, India, from 14 to 16, show that the high accuracy of the multiple regression forecasting model forecasts demonstrates the superior effectiveness of the model in differentiating it from time series forecasting. The causal analysis forecasting model presents the contradictory links in the development of rail freight and creates a model with a high degree of confidence, although the limitations of the model principle itself make it difficult to create a data model that can fully and objectively reflect the complex causal relationships inherent in the series.

In general, traditional time series forecasting and causal analysis forecasting methods, with their simplicity and low modeling difficulty, are highly suitable for situations where environmental factors are subject to small changes or significant linear relationships. With the rapid development of the economy and society, the traditional forecasting methods based on statistics gradually reveal various shortcomings, the emergence of modern intelligence and its highly developed for railway freight volume forecasting into the next technological era to build a new talent technology ladder.

3.2.2 Modern Forecasting based on Machine Intelligence

At present, rail freight forecasting requires a large amount of data to support, and with the complexity and variability of forecasting objects, relying on machine-intelligent technological forecasting is becoming the preferred solution for efficient problem-solving. Forecasting using model creation relies too heavily on data-based probabilistic extrapolation, whereas machine learning through data-driven approaches can minimize forecast errors. The statistically-based model prediction is based on the premise that models are 'defaulted' to common-sense laws, with unbiased base tests used to reduce prediction error, ignoring the inherent error between the model and the laws of change in reality. By automating the analysis of data variation and obtaining the laws through machine intelligence, the limitations of the model or assumptions are effectively circumvented, and the value of the forecast error in freight volumes is truly minimized.

The use of machine intelligence forecasting methods is commonly found in grey forecasting[7], support vector machine forecasting[8], neural network forecasting and other methods[9].

(1) Grey forecasting methods

The most typical feature of using the grey system for forecasting is that it uses data series based on the analysis of the laws of the original data series to generate new laws, suitable for short and medium-

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0081

term forecasting, especially in the case of insufficient information on the original data is better. Common models are the GM(1,1)[10], and the grey Verhulst model[11].

While using the grey forecasting method for forecasting, Markov chains were added to optimize the input variables. The results show that the accuracy of the grey forecasting model is greatly influenced by the magnitude of the initial data transformation and further optimization is required. The use of the grey Verhulst model for forecasting cargo volumes in civil aviation also confirms the effectiveness of the grey forecasting method when applied to long-term forecasting. Using the transport structure and capacity arrangement of Guangdong as a theoretical basis, it was found that the GM(1,1) model, which is suitable for high stability and regularity of the initial data series, and the grey Verhulst model, which is suitable for first-order non-linear dynamic models, are mainly used for sample series with small data volumes and simple calculations. Both models predict well under certain conditions, although the ability to generalize is much weaker.

As the grey prediction model is aimed at data series with low completeness and disorder, the prediction results are also mostly rough and abbreviated. Some researchers have tried to improve the effectiveness of the grey forecasting method by using the power model improvement idea to deal with the initial data series to create an unbiased grey Verhulst model. The accuracy of the improved grey prediction model did improve. The comparative analysis shows that the grey forecasting method still has excellent applicability in the face of low completeness, high volatility of the initial data series and the presence of many unexpected factors.

(2) Neural network forecasting method

Neural network forecasting is a non-linear dynamical forecasting model that relies on the structural functions of the human brain. The advantages of this model include parallel computing and distributed storage, as well as its adaptive and self-learning characteristics, which are widely used in several fields such as logistics and signal processing. Currently, the most commonly used neural network algorithms are BP neural networks (BPNN)[12, 13], grey neural networks (GNN)[14], radial basis function neural networks (RBFNN)[15], wavelet neural networks (WNN)[16] and long and short-term memory neural networks (LSTM)[17]. Based on the working principle of the neural network prediction method, the parallel mode, when making predictions, is prone to problems such as local optimality, especially in the face of a very small amount of initial data and the coexistence of noise, this shortcoming needs further optimization.

(3) Support vector machine prediction

Support vector machine (SVM) prediction is not only efficient but also has strong extrapolation capabilities and can be used in combination with other algorithms for prediction with small sample sizes, non-linearities or high-level pattern recognition types. The key to improving the efficiency of the input of high-dimensional data lies in the dimensionality reduction of the initial data series.

3.2.3 Portfolio Forecast

For sudden data fluctuations, a single forecasting model cannot respond in a timely and effective manner, but a combination of forecasting methods can bring together the strengths of all.

(1) Weight allocation type

A new model is created by assigning weights to two or more forecasting methods. For example, to forecast the volume of freight crossing the Three Gorges Hub from 19 to 22 years, a model is created using a combination of traditional grey forecasting models and neural network forecasting models, based on the ordered weighted geometric mean (IOWGA) assignment method, and the error in the forecast results is less than that of a single grey or neural network forecasting model.

(2) Data pre-processing

The initial data is processed using the difference method or wavelet analysis, and the processed data series are input into the prediction model[18]. From the point of view of data updating, new information can be fully utilized to improve the model's ability to deal with unexpected situations, and from the point of view of research ideas, this provides a new direction for the creation of dynamic, emergent data models for railway freight forecasting.

(3) Parameter and structure optimization

Optimized parameters and structures are selected and applied to the combined forecasting model[19, 20]. For example, CUI et al. used an improved wavelet grey forecasting model for regional rail freight forecasting, using wavelet analysis technology and inputting optimized and filtered variable data through grey correlation analysis, and the results showed that the combined forecasting model created converged quickly and with high forecasting accuracy[21].

(4) Error correction type

Using error correction to process the model prediction results can effectively improve the prediction accuracy. For example, the GM(1,1) model, the combined GM(1,1)-Fourier error correction (EFGM) model, and the Gray Verhulst model (EFGVM) are used to forecast short-term traffic speeds and travel times, and the error-corrected models are more accurate than the non-linear time series models[22-24].

The above forecasting methods are more suitable for in-depth study as improvements than combined forecasting methods. A combined forecasting model that combines the advantages of multiple single models can fully avoid the shortcomings of a single model and create a forecasting model with higher comprehensive forecasting and scientific analysis capabilities and stronger extrapolation power. The rapid socio-economic development has also intensified the emergence of more dynamic and unexpected factors. Therefore, creating a more optimal solution of the combined forecasting model lecturer is a main exploration direction in the future railway freight volume forecasting.

4. The Development of Railway Freight Forecasting

The above research results show that researchers have made further optimization and improvement of the railway freight forecasting methods based on traditional statistical model forecasting methods, combined with the application of new computer science and technology tools. In the face of the increasingly complex and variable structure within the railway freight volume, coupled with the emergence of various unexpected factors, various methods and models used in forecasting have emerged.

(1) Broadening the scope for improvement of a single type of forecasting model with the selection of better parameters and structural optimization. In improving grey forecasting models, better grey parameters can be chosen; in improving new network forecasting models, more experience is spent on the problem of determining the optimal structure; in support vector machine forecasting, the selection of better parameters and the optimization of the kernel function is key to improving the accuracy of the whole model.

(2) Forecasting with full consideration of massive, high-latitude and heterogeneous data is not only a current hotspot in the field, but also a difficult one. With the development of the closer relationship between economic structure and the logistics field, the raw data is increasingly showing the characteristics of massive, high latitude and heterogeneous. Traditional forecasting methods are no longer capable of dealing with such a huge amount of data, so the only way to build a forecasting model with the characteristics of big data, combining the strengths of traditional and modern intelligence, fully excavating and analyzing the various data factors related to railway freight transport to derive the laws of change, and making forecasts based on the laws is the key development direction and a hot and difficult problem in the future.

5. Conclusion

The development of railway freight forecasting has played an important role in optimizing and improving the top-level design and organizational management of railway logistics and freight transport. In this regard, our researchers have invested a lot of research effort in creating and improving various forecasting models, and have achieved a series of significant theoretical and practical results. With the ultimate goal of making forecasts "scientifically valid, reasonably accurate", single or component forecasting models are created for different environmental contexts using qualitative and quantitative or comprehensive analysis methods, and then the final forecasting results are derived through analysis and research, resulting in a "theoretical research - technical analysis - engineering application" process. --Technical analysis --Engineering application" integrated system.

To sum up, this paper mainly analyses and discusses the existing research results on railway freight forecasting, based on various forecasting methods and models known so far, analyses the various factors affecting the railway freight volume, studies the development status of railway freight forecasting, and outlooks the future development trend of railway freight forecasting accordingly, and briefly explains some potential research directions for the railway freight (1) The existing research results are usually not available.

(1) The existing research results usually focus on a single type or combination of forecasting models, which have a high accuracy of forecasting results in a short period, but with the change of the time range, the accuracy error will increase. Therefore, it is important that the study forecasting methods in different categories depending on the length of the time horizon, and create more efficient and accurate combination forecasting models for different time horizons.

(2) To consider the current social environment background, such as the new crown epidemic, earthquake or other extreme events, will have a disruptive impact on the prediction of the development trend of railway freight volume, to improve the level of emergency management of the railway freight system for unexpected factors, it is necessary to fully reflect this type of impact factors in the prediction model, and the short period of the railway freight volume change pattern prediction calculation, to effectively improve the accuracy and reliability of the prediction results. Effectively improve the accuracy and reliability of the forecast results.

(3) To combine with the current level of social and economic development, railway freight volumes are subject to cyclical fluctuations due to economic influences; different regions and freight types can cause seasonal fluctuations in freight volumes; extreme weather disasters, environmental policies change in time, social emergencies and other factors can have a significant impact on railway freight volume forecasting results. The existence of cyclical, seasonal and unexpected factors makes it difficult to forecast rail freight volumes accurately, but taking into account these factors to make more accurate forecasts of their development trends will greatly contribute to the improvement of the organization and planning of freight logistics in different regions.

Acknowledgments

Thanks to the reviewers for their valuable comments.

References

[1] L. Liu, Y. Zhang, C. Chen, et al. Is the spatial-temporal dependence model reliable for the short-term freight volume forecast of Inland ports? A case study of the Yangtze River, China, Journal of Marine Science and Engineering, vol. 9 (2021), 985-985.

[2] S.S. Wen, H.Z. Ye. Analysis on VAR model of demand structure for China's Railway Freight Transportation, Railway Transport And Economy, vol. 30 (2008), 1-5.

[3] S.B. Bai. Research on railway passenger and freight volume based on linear regression model, Western Transportation Technology, vol. 10 (2010), 141-144.

[4] L. Liu, A. Xie, H. Ping, et al. Research on freight development of Guangdong province based on grey theory model, Mathematical Problems in Engineering, vol. 2021 (2021), 1-14.

[5] O. Trull, J.C. Garcia-Diaz, A. Troncoso. Initialization methods for multiple seasonal Holt-Winters forecasting model, Mathematics Week, vol. 8 (2020), 1-16.

[6] Q.Q. Hui. Road freight volume forecasting and analysis based on ARMA model, Economic Research Guide, vol. 12 (2020), 35-37.

[7] Y.A. Wang, X.M. Chen, Y.H. Han, et al. Forecast of passenger and freight fraffic volume based on elasticity coefficient method and grey model, Procedia-Social and Behavioral Sciences, vol. 96 (2013), 136-147.

[8] Y. Liu, M.X Lang. Railway freight volume prediction based on support vector regression (SVR), Applied Mechanics and Materials, vol. 3308 (2014), 587-589.

[9] C. Yang, X.M. Li. Research on railway freight volume prediction based on neural network, E3S Web of Conferences, vol. 143 (2020), 01050.

[10] F. Yang, X. Tang, Y. Gan, et al. Forecast of freight volume in Xi'an based on Gray GM (1, 1) model and Markov forecasting model, Journal of Mathematics, vol. 2021 (2021), 1-6.

[11] B. Zeng, M. Tong, X. Ma. A new-structure grey Verhulst model: Development and performance comparison, Applied Mathematical Modelling, vol. 81 (2020), 522-537.

[12] C. Zhou, J.C. Tao. Adaptive combination forecasting model for China's logistics freight volume based on an improved PSO-BP neural network, Computers, Networks & Communications, vol. 44 (2015), 646-666.

[13] C. Zhong. Prediction of chinese port cargo volume based on BP algorithm, Journal of Physics: Conference Series, vol. 1982 (2021), 37-42.

[14] Q. Ke, P. Deng. Improved grey neural network model based on prediction of over-dam cargo volume at Three Gorges Hub, Jouenal Of Shanghai Maritime University, vol. 42 (2021), 82-87.

[15] N.B. Karayiannis, M.M. Randolph-Gips. On the construction and training of reformulated radial basis function neural networks, IEEE transactions on neural networks, vol. 14 (2003), 835.

[16] J.W. Liu, F.L. Zuo, Y.X. Guo, et al. Research on improved wavelet convolutional wavelet neural networks, Applied Intelligence, vol. 2020 (2020), 1-21.

[17] Y.Q. Yue, Y.Y. Qiu, Q. Zhang, et al. Forecasting the railway freight volume in China based on combined PSO-LSTM model, Journal of Physics: Conference Series, vol. 1651 (2020), 012029.

[18] W. Wang, J. Jin, Y. Li. Prediction of inflow at Three Gorges Dam in Yangtze River with wavelet network model, Water resources management, vol. 23 (2009), 2791-2803.

[19] Z. Guo, J. Fu. Prediction method of railway freight volume based on genetic algorithm improved general regression neural network, Intelligent Systems, vol. 28 (2019), 43-56.

[20] Y. Sun, M. Lang, D. Wang, et al. A PSO-GRNN model for railway freight volume prediction: empirical study from China, Journal of Industrial Engineering and Management, vol. 7 (2014), 413-433.

[21] N.D. Cui, W.L. Xiang, X.L. Meng, et al. Research on freight volume prediction based on wavelet grey GM(1,1) model, Journal of Railway Science and Engineering, vol. 14 (2017), 2480-2486.

[22] G. Comert, N. Begashaw, N. Huynh. Improved grey system models for predicting traffic parameters, Expert Systems With Applications, vol. 177 (2021), 114972.

[23] B. Anton, C. Gurcan. Short-term freeway traffic parameter prediction: Application of grey system theory models, Expert Systems With Applications, vol. 62 (2016), 284-292.

[24] C. Jie, H. Ma, C. Yuan, et al. Novel grey verhulst model and its prediction accuracy, Journal Of Grey System, vol. 27 (2015), 47-53.