

Research Article

ARIMA-FSVR Hybrid Method for High-Speed Railway Passenger Traffic Forecasting

Meng Ge , Zhang Junfeng , Wu Jinfei , Han Huiting , Shan Xinghua ,
and Wang Hongye 

Institute of Computing Technology, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China

Correspondence should be addressed to Wang Hongye; llxxff_mg@126.com

Received 12 March 2021; Revised 21 April 2021; Accepted 22 May 2021; Published 30 May 2021

Academic Editor: Chih-Cheng Hung

Copyright © 2021 Meng Ge et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the prediction accuracy of railway passenger traffic, an ARIMA model and FSVR are combined to propose a hybrid prediction method. The ARIMA prediction model is established based on the known railway passenger traffic data, and then, the ARIMA prediction results are used as the training set of the FSVR method. At the same time, the air price and historical passenger traffic data are introduced to predict the future passenger traffic, to realize the mixed prediction of railway passenger traffic. The case study demonstrates that the hybrid prediction method can effectively improve the prediction performance of railway passenger traffic. Compared with the single ARIMA method, the hybrid prediction method improves the delay of the prediction results. Compared with the FSVR prediction result, the hybrid prediction method greatly reduces the errors in the extreme points of passenger traffic and long-term prediction. The relevant research results of this paper provide a useful reference for the prediction of railway passenger traffic.

1. Introduction

At present, commonly used passenger flow prediction methods are based on historical data including time-series methods, support vector machines, and neural networks [1–3]. For instance, Ni et al. [4] applied the autoregressive moving average (ARIMA) method to solve traffic flow prediction and proved that it can solve the problem of modeling about nonstationary time-series prediction. Xie et al. [5] designed the fuzzy time-series ARIMA method for long-term waterway traffic volume prediction. Li et al. [6] proposed a robust v -support vector regression (RSVR) method to forecast vessel traffic flow. Liu et al. [7] adopted a support vector machine- (SVM-) based regression prediction to predict the bus passenger flow in the target time window. Li et al. [8] put forward a backpropagation neural network (BPNN) model with population per distance band for traffic flow prediction of urban rail transit station. Hu et al. [9] developed a model re-sample recurrent neural network (RRNN) to forecast passenger traffic on mass rapid transit systems.

Due to the different advantages and disadvantages of various prediction methods, the prediction effect of a single

mechanism prediction method is often not ideal. If two or more methods are organically combined to form a hybrid prediction method, it will overcome the deficiencies of a single prediction mechanism and improve the performance of passenger flow prediction [10, 11]. Khan et al. [12] combined wavelet transform (WT) with artificial neural network (ANN) and ARIMA into a hybrid model for meteorological drought forecasting, and the model inherits the merits of both WT and ANN-ARIMA. Wu et al. [13] created a hybrid model of ARIMA and wavelet neural network (WNN) combined with genetic algorithm to predict the river water quality. Yu et al. [14] built a novel SVR-ANN combined model with EEMD for rainfall prediction. Luo et al. [15] explored a combined prediction model based on the empirical mode decomposition, support vector regression, and wavelet neural network (EMD-SVR-WNN) to forecast the structural settlement and deformation. The above models achieved satisfactory results. It can be found that SVR and neural network are suitable for solving complex nonlinear problems, and the time-series model has great advantages for time-based prediction. However, there are still some inherent defects in the neural network model, such

as ease of sinking into local optimization and the overfitting. Therefore, the SVR and time-series method are selected for hybrid prediction.

In this thesis, a combination of differential integrated moving average autoregressive model (ARIMA) and fuzzy support vector regression machine (FSVR) is used to implement a mixed forecasting strategy for railway passenger flow. And, apply it to the actual passenger flow forecast of Shanghai-Guangzhou high-speed railway in order to obtain good forecast performance. Support vector regression (SVR) is a general learning method based on the statistical learning theory of limited samples (SLT) [16]. Fuzzy support vector regression (FSVR) is a new type of support vector regression machine that combines fuzzy mathematics and support vector regression. It introduces fuzzy membership and improves the generalization of machine learning ability. According to the theory of time-series analysis, the ARIMA model is suitable for the prediction and analysis of stationary time series, and the passenger flow data is generally non-stationary series, which needs to be smoothed by difference. Therefore, the differential autoregressive moving average model (ARIMA) is used to predict passenger flow.

2. ARIMA

Differential autoregressive moving average model (ARIMA) is an important method for studying time series. In ARIMA (p, d, q), AR is autoregressive and p is the number of autoregressive items, MA is the moving average, q is the moving average item number, and d is the number of differences made to make it a stationary sequence. The ARIMA (p, d, q) model is an extension of the ARMA (p, q) model.

The basic form of the ARMA model is

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad (1)$$

where c is the constant, $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ is the coefficient, ε_t is the white noise sequence, p is the autoregressive order, and q is the moving average order.

After passing the difference, the basic form of the ARIMA model is

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t, \quad (2)$$

where L is the lag operator and d is the difference order, $d \in \mathbb{Z}, d > 0$.

3. Fuzzy Support Vector Regression

The principle of FSVR is to find a function by minimizing the prediction error, use the nonlinear mapping function ϕ to map the data x_i in the input space to the high-dimensional space H , and perform linear regression calculation in H to achieve the effect of nonlinear regression in the original low-dimensional space [17].

In practical applications, different data points contribute differently to the training results, so FSVR solves the problem of

overlearning due to the presence of noisy data by introducing fuzzy parameters to eliminate the influence of noise [18], that is, there is a fuzzy degree and each data point is connected so that a training set with fuzzy members will be generated.

For FSVR, let the training set be $S = \{(x_1, y_1, s_1), (x_2, y_2, s_2), \dots, (x_N, y_N, s_N)\}$, where $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$, and $s_i \in [0, 1]$. In the time-series problem, the membership degree s_i is a function of the time series t_i ($1 \leq i \leq N$). In this thesis, the fuzzy membership function $f(t_i)$ is the quadratic function of the time series t_i , namely, $s_i = f(t_i)$:

$$s_i = (1 - \lambda) \left(\frac{t_i - t_1}{t_N - t_1} \right)^2 + \lambda. \quad (3)$$

The boundary conditions are

$$\begin{cases} s_1 = \lambda & (0 < \lambda \leq 1), \\ s_N = f(t_N) = 1. \end{cases} \quad (4)$$

FSVR is for solving quadratic programming problems:

$$\begin{aligned} \arg \min_{\omega, b, \xi, \xi^*} \phi_L(\omega, \xi, \xi^*) &= \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N s_i (\xi_i + \xi_i^*), \\ \text{s.t.} \begin{cases} y_i - \omega^T \phi(x_i) - b \leq \varepsilon + \xi_i, \\ \omega^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i \geq 0, \xi_i^* \geq 0, \end{cases} \end{aligned} \quad (5)$$

where ω is the regression hyperplane weight vector, b is the deviation coefficient, C is the penalty parameter (as a constant value), ε is the regression hyperplane bandwidth, ξ_i and ξ_i^* are the relaxation variable, and s_i is the fuzzy membership.

The dual form of equation (5):

$$\begin{aligned} \arg \max_{\alpha, \alpha^*} \phi_L(\alpha, \alpha^*) &= -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \\ &\quad - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*), \\ \text{s.t.} \begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, s_i C]. \end{cases} \end{aligned} \quad (6)$$

Solving the dual problem (6), we can get the FSVR regression function:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b. \quad (7)$$

4. Experiments

Using the high-speed rail passenger flow between Shanghai and Guangzhou as experimental data, the passenger flow is

obtained by day, a total of 176 days of sample data are collected, the first 165 days of sample data are used to build the model, and the last 8 days of sample data are used as test samples to predict comparative analysis.

In order to reduce the computational complexity and accuracy of parameter selection, the raw data is normalized. Table 1 shows part of the passenger flow data.

Using the ARIMA model to predict the values, the results are as follows.

It can be seen from the prediction results shown in Figure 1, and the ARIMA model can realize the prediction and analysis of railway passenger traffic. The fluctuation of its prediction results is consistent with the actual passenger traffic curve, but there is a large delay phenomenon which causes a large prediction error and the prediction effect is not ideal.

Based on FSVR's passenger flow prediction, the results are as follows.

It can be seen from the prediction results shown in Figure 2, and the FSVR has a strong nonlinear approximation ability; it has shown good prediction performance in the process of railway passenger traffic forecast, especially in the short-term passenger traffic forecast; its prediction error is small, and the passenger traffic continues to increase or continue. The prediction error is small during the decrease, but at the extreme point, where the passenger traffic trend changes, that is, the passenger traffic changes from increasing to decreasing, or from decreasing to increasing, the prediction error is large. In other words, the dramatic fluctuations in passenger traffic reduce the generalization ability of FSVR and affect its prediction performance.

Using the above ARIMA forecast results as the input items of FSVR, the mixed forecast of railway passenger traffic is realized. The results are as follows.

It can be seen from the prediction results shown in Figures 3 and 4 that the hybrid prediction method can combine the advantages of the two prediction methods to obtain the best prediction results. Compared with the ARIMA method, the delay of the hybrid method prediction results is greatly improved; compared with the FSVR, the prediction effect at the extreme point is significantly improved, and the prediction error is greatly reduced.

In order to prove the performance of the proposed algorithm, it is compared with the ARIMA-WNN method and the EMD-SVR-WNN method. The results of the three hybrid prediction methods are shown in Figure 5.

It can be seen from the prediction results in Figure 5 that, though the ARIMA-WNN method is accurate in the early prediction, it gradually appears the phenomenon of delay after 4 days. The overall trend of the EMD-SVR-WNN method is consistent with the original data; however, the overall predicted value is small. Compared with the above two methods, the prediction results of the ARIMA-FSVR method are more accurate. The forecast error indexes of various methods are shown in Table 2.

TABLE 1: Part of the passenger flow data.

SN	Date	Passenger flow	Normalization
1	20190517	1431	0.4630
2	20190518	1421	0.4475
3	20190519	1466	0.5170
4	20190520	1779	1.0000
5	20190524	1213	0.1265
6	20190525	1219	0.1358
7	20190526	1138	0.0108
8	20190527	1131	0.0000

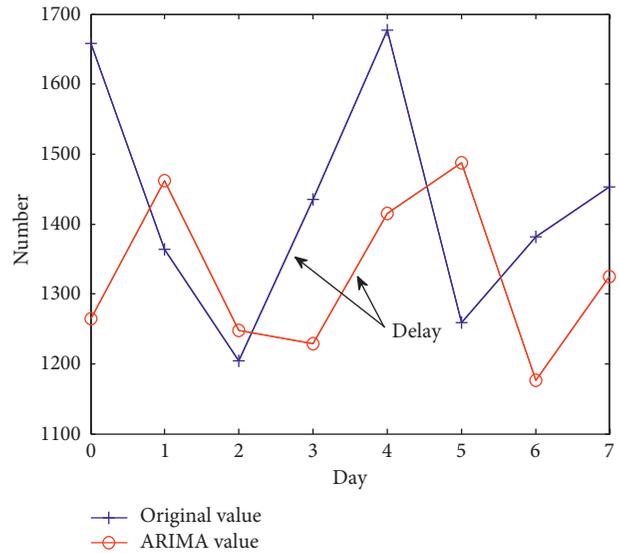


FIGURE 1: ARIMA results.

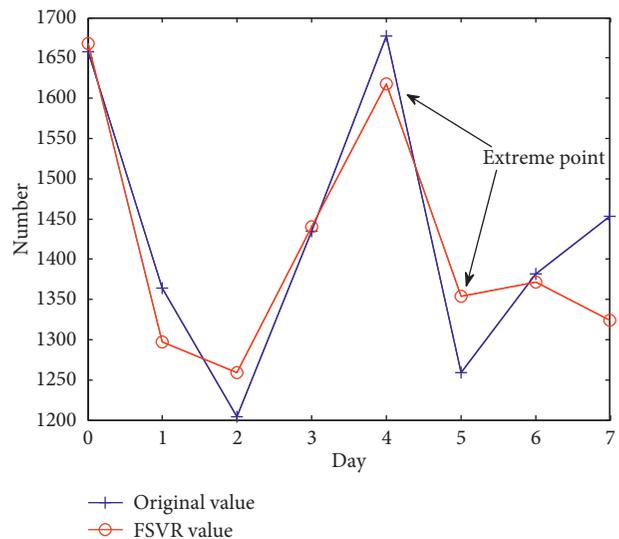


FIGURE 2: FSVR results.

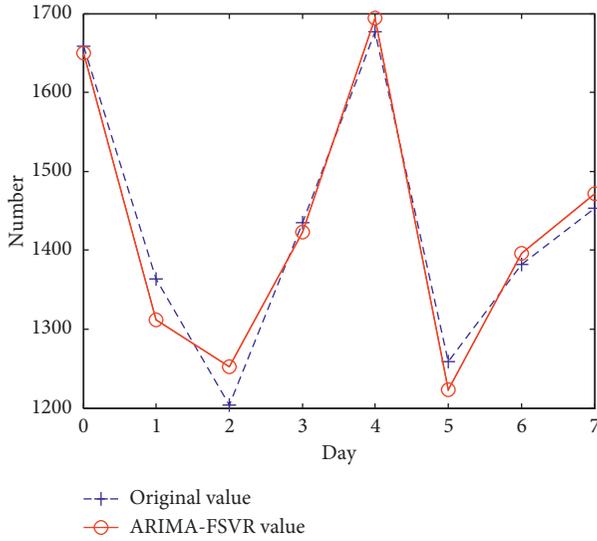


FIGURE 3: ARIMA-FSVR results.

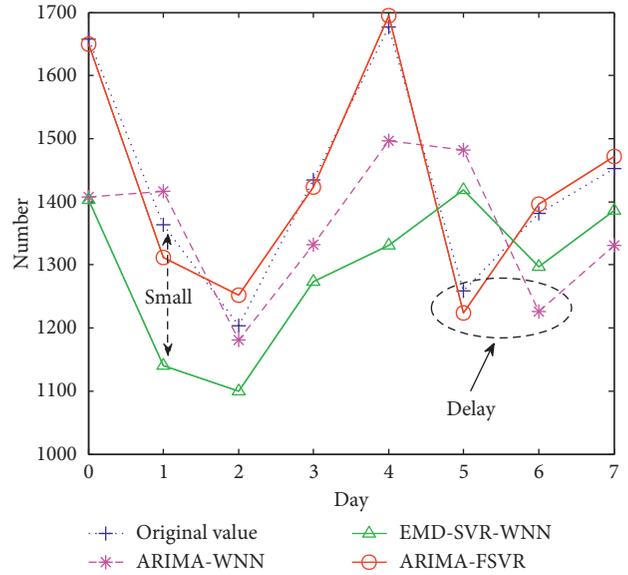


FIGURE 5: Three hybrid prediction results vs. original values.

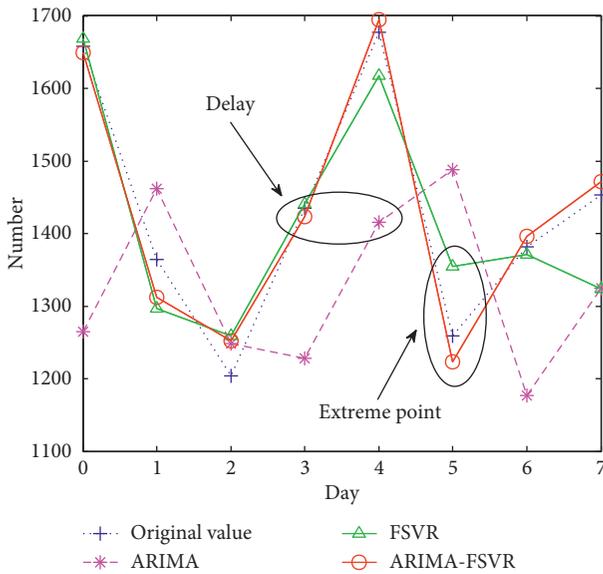


FIGURE 4: The ARIMA-FSVR, ARIMA, and FSVR predicted results vs. original values.

It can be seen from Table 2 that the standard error of the ARIMA-FSVR prediction is smaller than the ARIMA and FSVR methods. It is also smaller than the other two hybrid methods. The correlation coefficient of the ARIMA-FSVR method is less than 0.0001, and the P value is 0.9822. Compared with the other four methods, the correlation coefficient is larger and the P value is lower, which proves that the trend of the ARIMA-FSVR method is more accurate and can accurately predict the railway passenger traffic.

It can be found from the experimental results that the ARIMA-FSVR method can accurately predict the railway passenger traffic, handle complex nonlinear relationships, and obtain satisfactory prediction results.

TABLE 2: Forecast error indexes.

	ARIMA	FSVR	ARIMA-WNN	EMD-SVR-WNN	ARIMA-FSVR
RMSE	77.9183	24.0677	55.5990	69.5719	10.7347
Correlation coefficient	-0.0318	0.9065	0.4907	0.4849	0.9822
P value	0.9404	0.0019	0.2170	0.2233	0.0000

5. Conclusions

In this paper, a new hybrid method was successfully proposed which achieved great improvements regarding both the prediction accuracy and robustness of the single-item models:

- (1) The ARIMA-FSVR hybrid prediction method overcame the shortcomings exposed in the single-item forecasting method, and it can improve the ARIMA delay phenomenon.
- (2) The ARIMA-FSVR hybrid prediction method surmounts the extreme point problem of the FSVR method.
- (3) Empirical studies on the realistic passenger flow data indicated that the ARIMA-FSVR hybrid method was clearly superior to other benchmark hybrid models. This hybrid method obtained the lowest prediction error and had higher accuracy and more reliable prediction results.

In conclusion, the ARIMA-FSVR hybrid method can accurately predict the railway passenger traffic, overcoming the shortcomings of the single-item forecasting method and, at the same time, merging the advantages of single-item forecasting and improving the accuracy of the forecast. This method effectively solves the nonlinear problem of railway

traffic data and provides a new and effective method for the nonlinear prediction problem in practical applications.

Data Availability

The case analysis data used to support this study are available from the railway passenger transport department upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The project was supported by Science and Technology Research Project of Beijing-Shanghai High Speed Railway Co., Ltd. (Grant no. Beijing-Shanghai Scientific Research-2020-2), Scientific Research Projects of China Academy of Railway Sciences Co., Ltd. (Grant no. 2019YJ120), and Science and Technology Research and Development Plan of China Railway (Grant no. K2019X022).

References

- [1] G. Ren and J. Gao, "Comparison of NARNN and ARIMA models for short-term metro passenger flow forecasting," in *Proceedings of the 19th COTA International Conference of Transportation Professionals*, Nanjing, China, 2019.
- [2] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, 2015.
- [3] T.-H. Tsai, C.-K. Lee, and C.-H. Wei, "Neural network based temporal feature models for short-term railway passenger demand forecasting," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3728–3736, 2009.
- [4] L. Ni, X. Chen, and H. Qian, "ARIMA model for traffic flow prediction based on wavelet analysis," in *Proceedings of the 2nd International Conference on Information Science and Engineering IEEE*, pp. 1028–1031, Hangzhou, China, December 2010.
- [5] Y. Xie, P. Zhang, and Y. Chen, "A fuzzy ARIMA correction model for transport volume forecast," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6655102, 10 pages, 2021.
- [6] M.-W. Li, D.-F. Han, and W.-L. Wang, "Vessel traffic flow forecasting by RSVR with chaotic cloud simulated annealing genetic algorithm and KPCA," *Neurocomputing*, vol. 157, pp. 243–255, 2015.
- [7] W. Liu, Q. Tan, and W. Wu, "Forecast and early warning of regional bus passenger flow based on machine learning," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6625435, 11 pages, 2020.
- [8] J. Li, M. Yao, and Q. Fu, "Forecasting method for urban rail Transit ridership at station level using Back propagation neural network," *Discrete Dynamics in Nature and Society*, vol. 2016, Article ID 9527584, 9 pages, 2016.
- [9] R. Hu, Y. C. Chiu, C. W. Hsieh, T. H. Chang, and L. Liao, "Mass Rapid Transit system passenger traffic forecast using a Re-sample recurrent neural network," *Journal of Advanced Transportation*, vol. 2019, Article ID 8943291, 14 pages, 2019.
- [10] S. Li, X. Liu, and A. Lin, "Fractional frequency hybrid model based on EEMD for financial time series forecasting," *Communications in Nonlinear Science and Numerical Simulation*, vol. 89, 2020.
- [11] M. A. Jallal, A. González-Vidal, A. F. Skarmeta, S. Chabaa, and A. Zerouala, "A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting applied to energy consumption prediction," *Applied Energy*, vol. 268, 2020.
- [12] M. M. H. Khan, N. S. Muhammad, and A. El-Shafie, "Wavelet based hybrid ANN-ARIMA models for meteorological Drought forecasting," *Journal of Hydrology*, vol. 590, Article ID 125380, 2020.
- [13] J. Wu, Z. B. Li, L. Zhu, and C. Li, "Hybrid model of ARIMA model and GAWNN for dissolved oxygen content prediction," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 48, pp. 205–210, 2017.
- [14] X. Yu, G. Ling, L. He, S. Xia, and W. Wang, "A SVR-ANN combined model based on ensemble emd for rainfall prediction," *Applied Soft Computing*, vol. 73, 2018.
- [15] X. Luo, W. Gan, L. Wang, Y. Chen, and X. Meng, "A prediction model of structural settlement based on EMD-SVR-WNN," *Advances in Civil Engineering*, vol. 2020, no. 4, Article ID 8831965, 11 pages, 2020.
- [16] X. Luo, D. Li, and S. Zhang, "Traffic flow prediction during the holidays based on DFT and SVR," *Journal of Sensors*, vol. 2019, no. 10, Article ID 6461450, 10 pages, 2019.
- [17] P. Huang, C. Wen, Li. P. Fu, Q. Y. Peng, and Z. C. Li, "A hybrid model to improve the train running time prediction ability during high-speed railway disruptions," *Safety Science*, vol. 122, 2019.
- [18] T. Bahraini, S. Ghazi, and H. S. Yazdi, "Toward optimum fuzzy support vector machines using error distribution," *Engineering Applications of Artificial Intelligence*, vol. 90, 2020.