PASSENGER FLOW FORECASTING OF ZHENGZHOU METRO LINE ONE BASED ON AN IMPROVED WHALE OPTIMIZATION ALGORITHM

by

Dasi HE[a*], Yupeng GOU[a], Weizhi YU[b], Sanxian XIA[c], and Jie LAN[b]

[a] School of Energy and Environment, Zhongyuan University of Technology, Zhengzhou, China,
[b] Urban Rail and Underground Engineering Design and Research Institute, China Railway Siyuan Survey and Design Group Co., Ltd., Wuhan, China,
[c] Technical Management Department, Zhengzhou Rail Transit Co., Ltd., Zhengzhou, China

Original scientific paper
https://doi.org/10.2298/TSCI2403315H

With the continuous expansion of urban scale in China, the increase of passenger flow has brought great pressure to the urban public transport system. An accurate and timely prediction of the short-term passenger flow at each metro station is extremely important for the metro intelligent control system to make a timely decision. In this paper, based on the measured passenger flow data of Zijingshan station of Zhengzhou Metro Line 1, an improved whale optimization algorithm is proposed to predict the passenger flow on different time scales. The results show that the method has higher accuracy than the traditional least squares support vector machines algorithm. The paper opens a new window for nowcasting warning in the rush hour and long-period optimization of the public transport.

Key words: *metro passenger flow, empirical mode decomposition, improved whale optimization algorithm*

Introduction

With the rapid development of urbanization, urban traffic congestion is becoming more and more serious. Vigorous development of public transport is an effective way to face this intractable problem. As one of the main public transport vehicles in the city, the metro is widely favored due to its high punctuality rate and large traffic capacity. But the passenger flow of metro stations might be too saturated in the rush hour. In order to make vehicle scheduling and staffing plan more reasonable, and maximize resource utilization for convenient transportation, accurate short-term forecast of metro passenger flow is very needed.

At present, there are mainly three methods to forecast the metro short-term passenger flow, *i.e.*, the time series model, the intelligent prediction model based on machine learning, and the combined forecasting model.

The first method is suitable for a long-period prediction, Li *et al.* [1] used SARIMA model to predict the passenger flow of Guangzhou south station, Xiao Lan station and Zhuhai station with a relatively good accuracy for a long-time prediction. The SARIMA model was also used [2] to predict the passenger flow of Beijing metro station, and a good prediction was obtained.

* Corresponding author, e-mail: hedasi@zut.edu.cn

The second method is to apply the machine learning [3] to achieve the prediction. The machine learning is widely used to predict timely human health [4-7], and it can be also used for nowcasting warning for passenger flow in the rush hour. Wang [8] established the BP neural network prediction model by analyzing the passenger flow distribution characteristics of urban rail transit, and achieved good results. Gu *et al.* [9] used the bus IC card boarding data to build the back propagation neural network and radial basis function (RBF) neural network. The test results showed that the RBF neural network prediction effect was better.

The third method is to couple the neural network with the filtering technology. Due to the data are always stochastic with noise, so the noise removal [10, 11] is much needed before prediction. Li and Chen [12] used the Elan neural network to predict preliminarily the passenger flow, which was then corrected by Kalman filter, the results show that its forecasting accuracy is better than that of a single passenger flow forecasting algorithm.

In this article, the whale optimization algorithm (WOA) [13] is applied for accurate prediction of the passenger flow, which was a nature-inspired meta-heuristic optimization algorithm, mimicking the social behavior of humpback whales. Compared with other algorithm, *e.g.*, the butterfly optimization algorithm [14], the genetic algorithm [15], Chun-Hui He algorithm [16-18], the WOA is much effective for predicting the passenger flow.

Improved Wolf Optimization Algorithm

The WOA is a new swarm intelligence optimization algorithm proposed in 2016 [13]. Compared with other swarm optimization algorithms, *e.g.*, the grey WOA [19] and the bee colony optimization algorithm [20], the main difference is that it uses random or optimal search agents to simulate hunting behavior, and uses spiral to simulate the bubble net attack mechanism of humpback whales. Its advantages are simple operation, few parameters and strong ability to jump out of the local optimum. Li *et al.* [21] analyzed the performance of bat algorithm [22], gray WOA [23], and WOA, indicating that these three algorithms have fewer parameters and higher accuracy and computational efficiency. Mirjalili and Lewis [13] tested several mathematical optimization and engineering optimization problems, and compared with firefly algorithm (FA) [24] and fruit fly optimization algorithm (FOA) [25]. The results showed that WOA was competitive in avoiding local optimization.

The WOA algorithm imitates the predation behavior of humpback whales. The position of each whale represents a feasible solution. The predatory behavior trajectory of humpback whales is the update method of the corresponding feasible solution. It is mainly divided into three steps which are surrounding prey, hunting behavior and searching prey.

Step 1: Humpback whales surround the prey when hunting. In order to describe this behavior, Mirjalili proposed the following mathematical model [13]:

$$D = |CX^*(t) - X(t)| \tag{1}$$

$$X(t+1) = X^*(t) - AD \tag{2}$$

where t is the current iteration number, A – the convergence factor, C – the swing factor, $X^*(t)$ – the position of the leader whale, $X(t)$ – the current position of the individual, and D – the iteration distance of the current humpback whale position approaching the best position. Expressions for A and C are:

$$A = 2ar_1 - a \tag{3}$$

$$C = 2r_2 \tag{4}$$

where a is a parameter which decreases linearly from 2 to 0 in the search iteration process, and r_1 and r_2 – random numbers of [0,1].

Step 2: According to the hunting behavior of humpback whale, it swims to its prey in a spiral motion. So, the mathematical model of hunting behavior is:

$$X(t+1) = \begin{cases} X^*(t) - AD & p < 0.5 \\ X^*(t) + D'e^{bq}\cos(2\pi q) & p \geq 0.5 \end{cases} \tag{5}$$

where p is a random number of [0, 1], q – the constant of [0,1], b – the constant which is used to describe the spiral shape, and D' – the distance between each humpback whale and the current optimal position. The expression of D' is:

$$D' = |X^*(t) - X(t)| \tag{6}$$

Step 3: Humpback whales use their location to conduct random search. The specific process is:

$$D = |X_{\text{rand}}(t) - X(t)| \tag{7}$$

$$X(t+1) = X_{\text{rand}}(t) - AD \tag{8}$$

where $X_{\text{rand}}(t)$ is a random individual whale in the current population position.

Although WOA has achieved good results in many practical applications, the ordinary whale optimization algorithm is still prone to fall into local optimization. Therefore, in order to further improve the performance of WOA algorithm, an improved WOA algorithm (IWOA) is proposed by introducing a reverse learning strategy and an adaptive weighting factor.

In the process of whale search, the reverse learning strategy is introduced to optimize the current population. That is to generate the reverse population of this population, and then the two populations are selected to select the better population. This increases the diversity of the population, which is to better improve the speed of finding the global optimal solution and increase the efficiency. At the same time, the weight factor W is introduced into the whale update formula to enhance the search ability and improve the accuracy. For the two improved additions, it plays a great role in improving the performance of the whale algorithm.

Improved empirical mode decomposition algorithm

Empirical mode decomposition (EMD) is a decomposition method that can decompose non-linear and non-stationary data. This method does not need to set parameters manually, and adaptively decomposes the data into several intrinsic mode functions (IMF) and a residual (RES) according to the time scale of the data itself.

Ensemble empirical mode decomposition (EEMD) is a new adaptive signal processing method, which is widely used to extract signals from the data generated in the non-linear and non-stationary process of noise. The EEMD method adds white noise to the original signal, and then uses EMD method to process these new sequences, which solves the modal confusion defect of EMD. The steps of EEMD are:

- A new time series $y(t)$ is generated by adding the white noise which obeys the normal distribution to the original sequence $x(t)$.
- The EMD method is used to decompose the newly generated time series. The result is that the new series is composed of several IMF components and a residual.
- Steps 1 and 2 to are repeated add the new normal white noise sequence to the signal.
- The integrated average operation on all IMF components is carried out to obtain the IMF component of passenger flow time series.

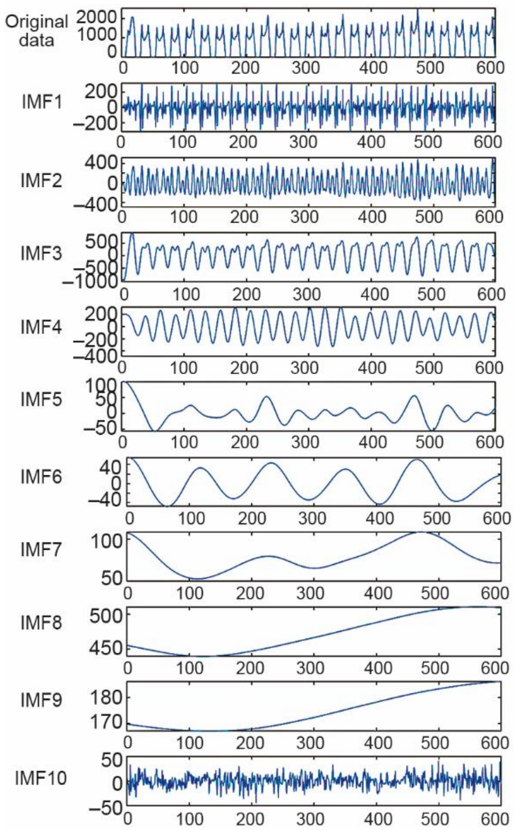

Figure 1. The EEMD decomposition results of passenger flow on weekdays

Due to non-linear and non-stationary of metro passenger flow data, it is not an easy to accurately model and predict the time series of metro passenger flow. The traditional forecasting methods are based on the assumption that these data are stationary signals which will inevitably lead to the loss of local transient information of some signals, which may affect the forecasting results of metro passenger flow. Many algorithms obtain the optimal solution through iteration. If the data contains a large number of noise data, it will greatly affect the convergence speed of the data, and even have a great effect on the accuracy of the training generation model. Therefore, the metro passenger flow data is decomposed into different components using EEMD, in order to reduce the interference of data noise. Taking 1 hour as the statistical interval, this paper constructs the time series metro passenger flow data of working days and non-working days respectively. The EEMD decomposition results are shown in fig. 1.

In fig. 1, the horizontal axis represents the length of data, and the vertical axis represents the size of data. The weekday data is decomposed into 10 IMF components and 1 residual. Each IMF component represents different time local characteristics in the data, and the residual represents the overall change trend of the data. IMF components are arranged according to the frequency from high to low. The first IMF component has the largest frequency, the largest amplitude and the shortest wavelength, while the frequencies of other IMF components decrease, the amplitudes decrease and the wavelengths increase in turn. The first few components represent the high-frequency period and noise of the original passenger flow data, reflecting the short-term change of the series, the last few components reflect the long-term change of the series, and the residual reflects the trend of the time series. The high-frequency components have low stationarity because they retain the non-stationarity of the original passenger flow data, while the low-frequency components do not contain the non-stationarity of the original passenger flow data and have high stationarity.

The EEMD-IWOA-LSSVM combined model

Based on the previous IWOA algorithm, EEMD technology and LSSVM model, a combined model is built in this paper. The idea is as follows.

– The original metro passenger flow $x(t)$ is decomposed by EEMD decomposition technology, and a series of subsequences are obtained.
– The IWOA-LSSVM prediction model is established for each sub sequence obtained by decomposition. The whale optimization algorithm is used to select the best penalty factor

and standardized parameters, radial basis function (RBF) as the kernel function, in LSSVM, and each sub sequence is trained and predicted.
– The final value is obtained by superimposing and summing the predicted values of each subsequence.

Prediction performance evaluation index

In order to evaluate the prediction performance of the algorithm, the following four performance index functions are used to evaluate the prediction performance.
– Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\overline{y_i} - y_i\right|^2} \tag{9}$$

– Mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\overline{y_i} - y_i\right| \tag{10}$$

– Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\overline{y_i} - y_i\right|}{y_i} \tag{11}$$

– Mean absolute percentage error (MAPE):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left|\overline{y_i} - y_i\right|^2}{\sum_{i=1}^{n}\left|Y_i - y_i\right|^2} \tag{12}$$

where $\overline{y_i}$ is the predicted value, y_i – the actual value, and Y – the mean value of sequence y_i.

Results and analysis

The original data in this paper comes from the card swiping data of all-in-one card of Zhengzhou Metro Line 1. Taking Zijingshan station as an example, the time range is from June 1, 2014 to July 5, 2014. A total of 480 groups of hour data in the first four weeks were selected as the training set, and a total of 120 groups of data in the fifth week were selected as the test set. The 192 groups of hourly data in the first 4 weeks were selected as the training set on non-working days, and 72 groups of data in the fifth week were selected as the test set.

Table 1 shows the prediction error analysis of working days, and tab. 2 shows the prediction error analysis of non-working days. Figure 2 shows the prediction results of different algorithms for working days at Zijingshan station. The whole week represents five working days in a week. Figure 3 shows the prediction results of different algorithms for non-working days at Zijingshan station. The whole week represents two non-working days in a week.

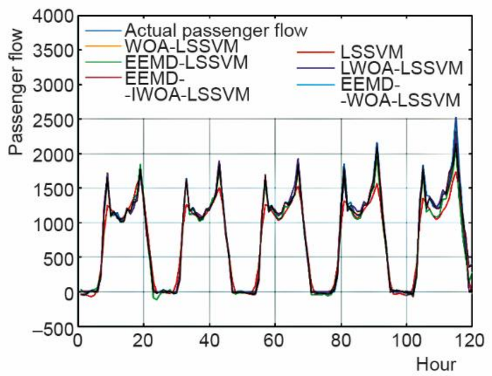

Figure 2. Forecast results of different algorithms for working days at Zijingshan station

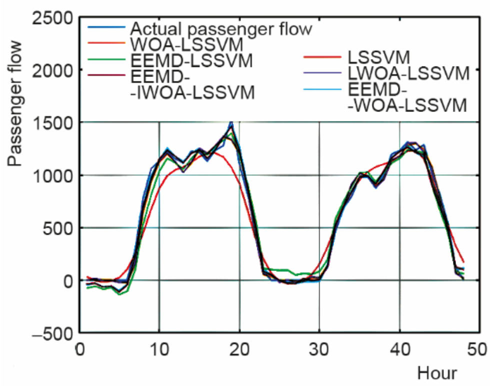

Figure 3. Forecast results of different algorithms for non-working days at Zijingshan station

Table 1. Comparison of forecast errors of passenger flow on weekdays

Prediction Model	RMSE	MSE	MAPE [%]	R^2
LSSVM	140.1125	90.5681	9.8554	0.95166
EEMD-LSSVM	124.0782	82.4306	8.9589	0.96186
WOA-LSSVM	78.8009	53.9572	5.6311	0.98287
IWOA-LSSVM	71.5261	51.6071	4.5887	0.98733
EEMD- WOA-LSSVM	59.4033	37.7712	3.5465	0.99126
EEMD- IWOA-LSSVM	59.0876	37.6617	3.5118	0.99135

Table 2. Comparison of passenger flow forecast errors on non-working days

Prediction Model	RMSE	MSE	MAPE [%]	R^2
LSSVM	147.7503	120.6406	11.1373	0.93896
EEMD-LSSVM	95.5778	78.2596	7.2081	0.97446
WOA-LSSVM	74.9899	53.9182	5.4835	0.98269
IWOA-LSSVM	68.6923	48.6875	4.8729	0.98681
EEMD- WOA-LSSVM	48.8595	37.4275	2.9113	0.99305
EEMD- WOA-LSSVM	47.6557	37.7225	2.9254	0.99365

In order to verify the prediction effect of the combined model, four groups of comparisons were made in this paper:

– Compared with EEMD-LSSVM and LSSVM models, the prediction accuracy of EEMD decomposition in weekday passenger flow forecast has been slightly improved, the fitting coefficient R^2 has increased from 0.95166 to 0.96186, and R^2 has significantly increased from 0.93896 to 0.97446 in non-weekday forecast. The difference of the forecast results of the decomposition method on different samples shows that these traditional passenger flow decomposition methods are not stable in the actual metro passenger flow forecast.

– Compared with IWOA-LSSVM and LSSVM models, the prediction accuracy has been significantly improved in both working and non-working days. The RMSE decreased from 124.0782 to 71.5261 on weekdays and from 147.7503 to 68.6923 on non-weekdays.
– Compared with WOA-LSSVM and IWOA-LSSVM models, RMSE decreased from 78.8009 to 71.5261 in weekday prediction, which was significantly improved. In non-weekday prediction, RMSE decreased from 74.9899 to 68.6923, indicating that the improved whale optimization algorithm has higher prediction accuracy than the traditional whale optimization algorithm.
– Compared with EEMD-IWOA-LSSVM model and other models, RMSE finally decreased to 59.0876 and the fitting coefficient increased to 0.99135 in the prediction of weekdays, while RMSE decreased to 47.6557 and the fitting coefficient increased to 0.99365 in the prediction of non-weekdays. The results show that the combined model of EEMD-IWOA-LSSVM plays an important role in improving the decomposition efficiency of metro passenger flow and reducing the random noise component in passenger flow, and realize the effective separation of passenger flow data and the privilege of characteristic information, which is conducive to in-depth analysis of passenger flow rules and passenger flow prediction.

Conclusion

In this paper, the data of metro passenger flow is mined, and the EEMD-IWOA-LSSVM combined forecasting model is designed according to the characteristics of daily passenger flow variation for a single LSSVM prediction model. The EEMD-IWOA-LSSVM combined model has better prediction accuracy and better applicability. Finally, this article only considers the workday classification of data and the predictive effect of the combined model, without considering external factors such as weather and holidays. In the future, it will be explored that different classifications and optimization algorithms can be carried out around external factors, cross-sectional passenger flow, commuting routes, *etc.*, and further improvements can be made to further improve the predictive effect of passenger flow, which is of greater significance for statistical research.

References

[1] Li, J., *et al.*, Passenger Flow Forecast of Guangzhou Zhuhai Intercity Railway Based on SARIMA Model (in Chinese), *Journal of Southwest Jiaotong University, 55* (2020), 1, pp. 41-51
[2] Zhao, P., Li, L., Research on Prediction of Urban Rail Transit Arrival Volume Based on ARIMA Model (in Chinese), *Journal of Chongqing Jiaotong University* (*Natural Science Edition*), *39* (2020), 1, pp. 40-44
[3] Mahmood, L., *et al.*, Assessment and Performance Analysis of Machine Learning Techniques for Gas Sensing E-nose Systems, *Facta Universitatis Series: Mechanical Engineering, 20* (2022), 3, pp. 479-501
[4] Wang, S. Q., *et al.*, Skeletal Maturity Recognition Using a Fully Automated System with Convolutional Neural Networks, *IEEE Access*, *6* (2018), July, pp. 29979-29993
[5] Wang, S. Q., *et al.*, An Ensemble-Based Densely-Connected Deep Learning System for Assessment of Skeletal Maturity, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52* (2020), 1, pp. 426-437
[6] Wu, K., *et al.*, 3D Convolutional Neural Network for Regional Precipitation Nowcasting (in Chinese), *Journal of Image and Signal Processing, 7* (2018), 4, pp. 200-212
[7] Wang, S. Q., *et al.*, Prediction of Myelopathic Level in Cervical Spondylosis Myelopathy Using Diffusion Tensor Imaging, *Journal of Magnetic Resonance Imaging, 41* (2015), 6, pp. 1682-1688

[8] Wang, L. N., Research on Passenger Flow Forecasting and Scheduling of Urban Rail Transit Network Based on Historical Passenger Flow Data (in Chinese), M. Sc. thesis., Jiaotong University, Beijing, 2011
[9] Gu, J. J., Jiang, Z. B., Bus Timetable Optimization in Rush Hour Based on Neural Network Passenger Flow Prediction (in Chinese), *Traffic Information and Safety*, *35* (2017), 2, pp. 109-114
[10] Gong, C. M., *et al.*, Tropical Algebra for Noise Removal and Optimal Control, *Journal of Low Frequency Noise, Vibration & Active Control*, *42* (2023), 1, pp. 317-324
[11] Wang, J., Tropical Algebra with High-Order Matrix for Multiple-Noise Removal, *Journal of Low Frequency Noise, Vibration & Active Control*, *42* (2023), 2, pp. 898-910
[12] Li, S. W., Chen, Y. S., Design and Simulation of Passenger Flow Forecasting Algorithm for Urban Rail Transit (in Chinese), *Computer Science, 41* (2014), 2, pp. 276-279
[13] Mirjalili, S., Lewis, A., The Whale Optimization Algorithm, *Advances in Engineering Software, 95* (2016), May, pp. 51-67
[14] Ghalambaz, M., *et al.*, Building Energy Optimization Using Butterfly Optimization Algorithm, *Thermal Science*, *26* (2022), 5A, pp. 3975-3986
[15] Jiang, M. X., Pan, Z. L., Optimization of Open Micro-Channel Heat Sink with Pin Fins by Multi-Objective Genetic Algorithm, *Thermal Science*, *26* (2022), 4B, pp. 3653-3665
[16] Khan, W. A., Numerical Simulation of Chun-Hui He's Iteration Method with Applications in Engineering, *International Journal of Numerical Methods for Heat & Fluid Flow*, *32* (2022), 3, pp. 944-955
[17] Khan, W. A., *et al.*, Numerical and Theoretical Investigation to Estimate Darcy Friction Factor in Water Network Problem Based on Modified Chun-Hui He's Algorithm and Applications, *Mathematical Problems in Engineering*, *2022* (2022), 8116282
[18] He, C. H., An Introduction to an Ancient Chinese Algorithm and Its Modification, *International Journal of Numerical Methods for Heat & Fluid Flow*, *26* (2016), 8, pp. 2486-2491
[19] Mirjalili, S., *et al.*, Grey Wolf Optimizer, *Advances in Engineering Software, 69* (2014), Mar., pp. 46-61
[20] Karaboga, D., Basturk, B. A Powerful And Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm, *Journal of Global Optimization, 39* (2007), 3, pp. 459-471
[21] Li, Y. L., *et al.*, Comparative Study of Several New Swarm Intelligence Optimization Algorithms (in Chinese), *Computer Engineering and Applications, 56* (2020), 22, pp. 1-12
[22] Li, C., A New Meta Heuristic Bat Algorithm, *Computer Knowledge and Technology, 6* (2010), 23, pp. 6569-6572
[23] Wang, Z. S., Pan T. H., A Gray Wolf Optimization Algorithm Using Position Mutation Strategy, *Proceedings,* 41st China Control Conference, Hefei, China, 2022, Vol. 4, pp. 208-211
[24] Yu, F. Z., Pan, D. Z., Improved Firefly Algorithm for Multi Robot Path Planning (in Chinese), *Computer Application Research*, *40* (2023), 3, pp. 1-6
[25] He, C., *et al.* An Improved Fruit Fly Optimization Algorithm and Its Application in Wet Flow Gas Desulfurization System, *Proceedings,* 32nd China Control and Decision-making Conference, Hefei, China, 2020, Vol. 2, pp. 891-896
[26] Zhang, J. Q., *et al.* Cloud Computing Task Scheduling Based on Orthogonal Adaptive Whale Optimization (in Chinese), *Computer Applications, 42* (2022), 5, pp. 1516-1523

Paper submitted: May 2, 2023
Paper revised: August 5, 2023
Paper accepted: August 7, 2023