Short-Term Passenger Flow Prediction Based on a Hybrid Model of Support Vector Machine and Long Short-Term Memory Network

Yuting Tian, Jinbing Ha*

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

* Corresponding Author Email: 1983318083@qq.com

Abstract. This study focuses on Line 10 of the Nanjing subway system, analyzing the spatiotemporal distribution characteristics of passenger flow data from the perspectives of network layout and station functionality. The research summarizes the unevenness, temporal variability, periodicity, and similarity characteristics of passenger flows along this line. Addressing these characteristics, a preprocessing method combining Self-Organizing Map (SOM) algorithm with Isolation Forest is proposed for handling passenger flow data, which involves data classification and anomaly detection. Input features such as periodic flows, weekday identification, weather conditions, and passenger flow categories are constructed. A Short-Term Passenger Flow Prediction Model (PSO-SVM) is developed, employing Support Vector Machine (SVM) algorithm optimized by Particle Swarm Optimization (PSO) for enhanced parameter tuning capability. The integrated predictive model is applied within the operational decision-making framework of Line 10, facilitating data storage, retrieval, and analysis of future station passenger flows. This enables rail transit operators to adjust operational strategies more accurately based on predicted passenger flows, thereby improving overall energy efficiency.

Keywords: Urban rail transit, passenger flow prediction, passenger flow analysis, support vector machine, long short-term memory neural network, BP neural network.

1. Introduction

With the vigorous development of China's economy and the accelerated urbanization process, the urban resident population continues to grow steadily, leading to increasingly severe traffic congestion issues.[1-3] In order to improve the traffic situation, China has invested substantial resources in urban rail transit construction. Urban rail transit boasts advantages such as high passenger capacity, high punctuality, energy conservation, emissions reduction, and comfort and convenience, particularly demonstrating significant characteristics of a low-carbon economy in ecological conservation. With increasing societal concern over climate change and energy consumption issues, the country has explicitly outlined in the 14th Five-Year Plan the promotion of comprehensive green and low-carbon transformation, aiming to achieve peak carbon dioxide emissions and carbon neutrality goals.[4] Consequently, urban rail transit plays an indispensable role in the public transportation system, increasingly favored by residents as their primary choice for daily travel.[5] However, as urban rail transit networks continue to expand and extend, facilitating daily travel for people, they also bring about some significant issues that cannot be ignored:

(1) Operating authorities find it challenging to adjust operational strategies based on real-time changes in passenger flow, leading to increased operational safety risks.

(2) The increase in passenger volume results in corresponding increases in energy consumption for rail transit, thereby reducing energy utilization efficiency.

Passenger flow forecasting involves the use of various appropriate theoretical knowledge and technical methods, considering relevant influencing factors, to analyze known historical data and information.[6] This allows for scientifically predicting future passenger numbers and demands. Passenger flow forecasting can be divided into long-term, medium-term, short-term, and very short-term forecasts based on time granularity.[7] Long-term forecasts cover periods exceeding one year into the future, while medium-term forecasts extend beyond one month, providing reference for the planning and construction of passenger transport. Short-term forecasts cover periods beyond one week but within one month, and very short-term forecasts cover periods within the next day, both serving as scientific guidance for train scheduling (timetables and operating plans).[8] Short-term passenger flow forecasting methods mainly include statistical, nonlinear, and neural network methods. Statistical models are primarily based on the assumption that future changes in passenger flow are similar to known historical patterns, using probability theory and mathematical statistics to establish models that precisely express historical passenger flow changes and infer future flows. Representative models include time series models and regression analysis models. Due to the nonlinear and volatile nature of rail transit passenger data, these forecasting models may not effectively handle such data, resulting in less precise predictions. Therefore, nonlinear prediction models and neural network-based prediction models are increasingly being introduced into passenger flow forecasting, including support vector machines, K-nearest neighbors, and random forests. For instance, Zhang and Xie constructed a v-support vector machine prediction model based on support vector machine theory, comparing its results with those of a multilayer perceptron neural network prediction model, and found their prediction model to be more effective.[9] Zhang Kai successfully combined the knowledge of least squares support vector machines (LSSVM) with wavelet kernel functions to establish the LSWSVM prediction model, further developing CPSO-LSWSVM and MCPSO-LSWSVM models through particle swarm optimization algorithms, providing effective prediction methods for both daily and sudden large passenger flows.[10] Guo Wen improved the support vector machine prediction model to a certain extent by using genetic algorithms to determine various parameters in the GA-SVM model.[11] Zhu Bingjie introduced the K-nearest neighbors algorithm into passenger flow forecasting, optimizing it with dynamic K-value methods and continually adjusting the error correction model, ultimately achieving more accurate predictions compared to the original K-nearest neighbor method.[12] Finally, Cha Min constructed a random forest model on the Spark engine, optimizing algorithms for calculating random forest similarity and selecting features, thereby reducing errors from minor influencing factors and enhancing prediction accuracy.[13]

This study is based on passenger flow data from Nanjing Metro Line 10, focusing on urban rail transit passenger characteristics. By integrating factors such as periodic passenger flow, weather indicators, weekday status, and passenger classification, a combined optimization forecasting model based on PSO-SVM and LSTM is designed and constructed. The model adjusts weight allocation appropriately to obtain forecasting results, thereby meeting operational demands of the metro enterprise. Experimental validation demonstrates that this combined model exhibits lower error metrics and higher prediction accuracy (89% and 87% accuracy for entry and exit forecasts, respectively). Successful application of this model facilitates proactive management of passenger flow data, enabling urban transit enterprises to adjust subway operation strategies more timely and accurately, thereby enhancing passenger convenience, improving energy efficiency, and accelerating the process of intelligent urban development.

2. Methodology

2.1. Prediction Model Based on Improved Support Vector Machine

Support Vector Machine (SVM) is an algorithm based on VC dimension theory and the principle of structural risk minimization.[14-16] The core idea is to construct an optimal hyperplane in the feature space to separate training samples, where the process of seeking the optimal hyperplane aims to maximize the margin. "Support vectors" are the specific sample points closest to this decision hyperplane. Under non-linear conditions, the training samples $\{(xj,\ yj)|j = 1,2,\ \dots\ ,\ s\}$ are mapped non-linearly to $G(xj,\ yj)$ in a higher-dimensional space, transforming the objective function (1) through similarity transformation.

$$\begin{cases} \min \frac{1}{2}\omega^T\omega + C\sum_{j=1}^{s}\theta_j \\ \text{s.t } y_j[\omega^T G(x_j+b)] \geq 1-\theta_j, \theta_j \geq 0 \end{cases} \tag{1}$$

By formulating the dual problem, the resultant regression model obtained is:

$$f(x) = \omega^T G(x) + b = \sum_{j=1}^{s} a_j^* G(x_j \cdot y_j) + b \tag{2}$$

In this context, G, also known as the kernel function, C represents the penalty parameter, and θ denotes the training error.

For optimization algorithms, we opt for Particle Swarm Optimization (PSO), chosen for its fast search speed and high execution accuracy. PSO simulates birds foraging where each particle iteratively updates its velocity and position under the guidance of a leading particle, converging from a dispersed state towards the target, i.e., the optimal solution of the problem.[17] The core of this algorithm lies in the updates of velocity vector $V(d)$ and position vector $X(d)$ for each particle. Additionally, the quality of the solutions found is evaluated through fitness. Specifically, during particle optimization, both individual best $p_{best,d}$ *and global best* $g_{best,d}$ solutions are recorded and particles continuously update their positions by following these two solutions. Based on this setup, calculations proceed as shown in Equation (3), where d denotes the iteration count, ω represents inertia weight, c denotes learning factor, and r signifies a uniformly distributed random number between [0,1].

$$\begin{cases} V_{d+1} = \omega \cdot V_d + c_1 \cdot r_1 \cdot (p_{best,d} - X_d) + c_2 \cdot r_2 \cdot (g_{best,d} - X_d) \\ X_{d+1} = X_d + V_{d+1} \end{cases} \tag{3}$$

In response to the issues of parameter dependency and weak fault tolerance in SVM, employing PSO for SVM optimization addresses both the complex nonlinear features of passenger flow data and utilizes a global heuristic optimization algorithm to seek the optimal solution for model parameters[5, 18]. The implementation of this model involves preprocessing the passenger flow data, inputting it into the model for training, and subsequently testing it with test data, followed by result analysis and comparison. The core structure primarily involves optimizing the penalty parameter and kernel parameter of SVM through PSO, where each particle represents an SVM model corresponding to its respective optimized parameters. The main steps are outlined as follows:

Step 1: Read and normalize processed sample datasets, dividing them into training and test sets.

Step 2: Select a kernel function and map the parameters to be optimized (penalty parameter (C), kernel parameter (γ)) to the particle swarm.

Step 3: Initialize the population, including population size, iteration count, initial particle velocities and positions, inertia factor, learning factor, acceleration constants, etc. The population size affects algorithm speed and diversity; typically recommended values range from 20 to 100. Inertia factor, ranging from 0.5 to 1, influences global and local optimization performance. Iteration count, typically between 50 to 600, balances solution stability and computational efficiency. Particle velocity limits are set according to SVM parameter ranges to prevent over speeding past optimal solutions or convergence difficulties. Each particle is represented by (C, γ).

Step 4: Define the fitness function to compute each particle's fitness value, determining the best position for each particle and the entire population. Update particle velocities and positions accordingly.

Step 5: Evaluate fitness values for each particle's new position, compare them with individual and population historical best positions, and decide whether to replace current optimal values based on newfound superiority.

Step 6: Determine termination conditions based on reaching the iteration limit or meeting a fitness threshold. If not met, iterate back to Step 4.

Step 7: Integrate the optimal (C, γ) values from Step 5 into the SVM prediction model for further training and evaluate its performance using test set data. If the set threshold is met, output final results; otherwise, continue optimizing relevant parameters.

2.2. Prediction Model Based on Long Short Term Memory Network

The Recurrent Neural Network (RNN) is an artificial neural network model that leverages its inherent recursive properties to process information characterized by strong spatiotemporal features or sequential dependencies.[19, 20] **Fig. 1** illustrates the basic structure of an RNN, with the detailed structure shown on the right side. Similar to traditional neural network architectures, the RNN introduces an additional layer of self-recurrence and updates its parameters through backpropagation. In this framework, x_t denotes the sequentially input data from time step 1 to t; h_t represents the output after activation at time step t; and S_t refers to the hidden state at time step t, which serves as the memory unit of the network. The final output h_t is generated by combining the output from the previous time step h_{t-1} and the current input data x_t, expressed mathematically in equation (4), where f is typically a nonlinear activation function. The matrices U, W, and V denote the weight parameters for the input data, the memory unit, and the output data, respectively.

$$St = f(Uxt + WSt{-}1) \tag{4}$$

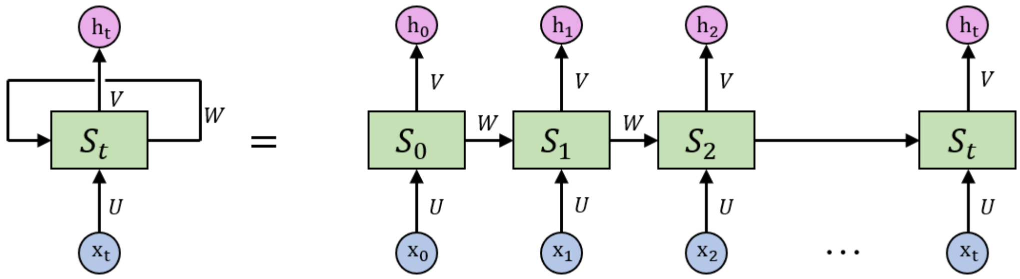


Fig. 1 Basic structure diagram of RNN

The Long Short-Term Memory (LSTM) network prediction model extends upon the foundation of Recurrent Neural Networks (RNNs) by incorporating forget and memory cells, effectively mitigating the issues of gradient vanishing and long-term dependencies. LSTM networks excel in short-term forecasting tasks.[21] The overall architecture comprises an input layer, hidden layers, fully connected layers, and an output layer. Initially, pre-processed passenger flow data is inputted through the input layer. Subsequently, the hidden layers propagate the passenger flow data, extracting relevant information features. This information is then transformed in dimensionality by the fully connected layers to retain useful information. Finally, the output layer generates predictions. The optimization process unfolds as follows:

Step 1: The prepared dataset is partitioned into training and testing sets. The model reads the passenger flow sequence $X = (x1, x2 \ldots xT)$, where T denotes the sequence length, and xi represents embedded vectors composed of confirmed influencing factors in this study, serving as model inputs at time i. Training parameters are initialized accordingly.

Step 2: Upon entering the LSTM layer, forward propagation computes output values. This primarily involves transmitting hidden state hi, where the entire layer shares network parameters $\omega 1$ and $b1$. Additionally, candidate values for memory state Ct are propagated.

Step 3: By employing an appropriate loss function, the model calculates error values for each neuron during backpropagation. This process involves backward propagation not only through the network layers but also along the time axis, thereby determining error terms for each neuron.

Step 4: Based on the training errors computed in Step 3, gradients for each parameter are determined.

Step 5: Using the optimization algorithm selected in Step 4, weights and other parameters are updated iteratively until convergence.

2.3. Combination prediction model based on weight allocation

To avoid the deficiencies of single-point algorithms and further enhance prediction accuracy, a composite forecasting model was designed, namely the Weight Allocation Model based on PSO-SVM and LSTM. The main concept of this model framework is as follows: after separately forecasting passenger flow using PSO-SVM and LSTM and considering additional input features (such as weather, weekday characteristics, passenger flow classification status), appropriate weights are adapted to the forecast results. This forms a pattern such as $W1 * result1 + W2 * result2 + W3 * result3$, resulting in the final prediction. This approach not only leverages the strengths of individual models but also complements them, and macroscopically adjusts and predicts the relevant influencing factors again, leading to more accurate predictions. [22, 23]

The structure of the Error Back Propagation (BP) algorithm is feedforward, composed of input layer, hidden layers, and output layer.[24] Its core involves backward feedback, where the output obtained by forward propagation between neurons is compared with the actual value to compute an error value. This error value is then propagated backward, automatically updating corresponding weights and thresholds iteratively until the desired error value is achieved.

Let f denote the activation function, μ denote the learning rate, ω denote weights, b denote thresholds, e denote error, and h denote hidden layer output. As shown in Equation (5), each iteration adjusts ω and b of each node to minimize e as much as possible:

$$\begin{cases} e = \dfrac{1}{2}\sum_{a=1}^{k}(y_a^i - x_a^k)^2 \\ \omega = \omega - \mu\dfrac{\partial e}{\partial \omega} \\ b = b - \mu\dfrac{\partial e}{\partial b} \end{cases} \tag{5}$$

The key step of the BP algorithm is to find the partial derivative solution of the above equation, as shown in equation (6):

$$\begin{cases} \mu\dfrac{\partial e}{\partial \omega} = -(y_a^i - x_a^k)f(x)' \text{ rvert } x_k(i)h_a^{i-1} \\ \mu\dfrac{\partial e}{\partial b} = -(y_a^i - x_a^k)f(x)'x_k(i) \end{cases} \tag{6}$$

Due to its high precision and strong adaptive capability, the BP neural network can effectively allocate weights among various inputs through its robust nonlinear mapping capability. In this study, the BP neural network is employed for the purpose of distributing combined weights among various inputs. The specific training steps of the composite forecasting model are outlined as follows:

Step 1: Prepare processed sequences of passenger flow data samples (x, y) across t groups, including predictions from PSO-SVM and LSTM, indicators for weekday features, weather factors, and passenger flow classification values. Specify the number of nodes for input and output neurons as m and n respectively, and set the number of nodes in the hidden layer to q.

Step 2: Compute the values of hidden layer neurons P based on input values, weights ω between neurons, and thresholds b. This computation is performed using Equation (7), where f denotes the activation function.

$$P_k = f\left(\sum_{i=1}^{m}\omega_{ik} - b_k\right) \tag{7}$$

Step 3: The output layer calculates the final output value Ot based on the hidden layer results and weights, using the formula (8).

$$O_t = \sum_{k=1}^{n} \omega_k \omega_{kt} - b_t \tag{8}$$

Step 4: Compare the output value of the previous step with the actual value to calculate the error result e, using the formula (9).

$$e = Y_t - O_t \tag{9}$$

Step 5: Backpropagate the calculated error and automatically update the weights and thresholds, as shown in equation (10), where μ is the learning rate.

$$\omega_{kt} = \omega_{kt} + \mu P_k e_t \tag{10}$$

Step 6: Continuously iterate the training until the preset error is met, otherwise proceed to Step 2 for continuous training.

3. Experiment results and analysis

3.1. Experimental Data Processing

This section primarily utilizes the passenger flow data of Yuantong Station on Nanjing Metro Line 10 from July 1, 2019, to November 24, with a time granularity of 60 minutes from 06:00:00 to 24:00:00, as the subject of simulation and prediction research. Specifically, the data from November 22 to 24 are designated as the test set for final evaluation of the model's predictive performance.

To effectively assess the composite model's performance, three commonly used evaluation metrics are employed: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). Their respective formulas are provided below:

$$MAE = \frac{1}{K} \sum_{i=1}^{K} |y_i - \widehat{y_i}| \tag{11}$$

$$MAPE = \frac{1}{K} \sum_{i=1}^{K} \left| \frac{y_i - \widehat{y_i}}{y_i} \right| \tag{12}$$

Among them, K represents the number of samples in the test set, yi is the actual passenger flow value, and $\hat{yi}$ is the predicted passenger flow value. The smaller the error values of these three evaluation indicators, the closer the actual passenger flow value is to the predicted passenger flow value, which means that the accuracy of the prediction model is higher.

Regarding the determination of input features, firstly, SOM is used for preliminary classification processing. Then, the passenger flow dataset is input into an isolated forest model to detect outliers and correct them. Finally, the visualized passenger flow trend of Yuantong Station is obtained as shown in **Fig. 2**, which conforms to the changing characteristics of a 7-day cycle.

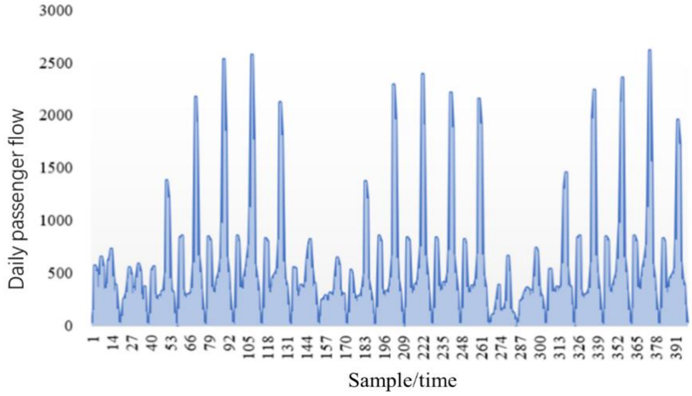


Fig. 2 Visualized passenger flow trend at Yuantong Station

Further perform SOM classification processing to obtain 5 types of passenger flow as shown in **Fig. 3**.

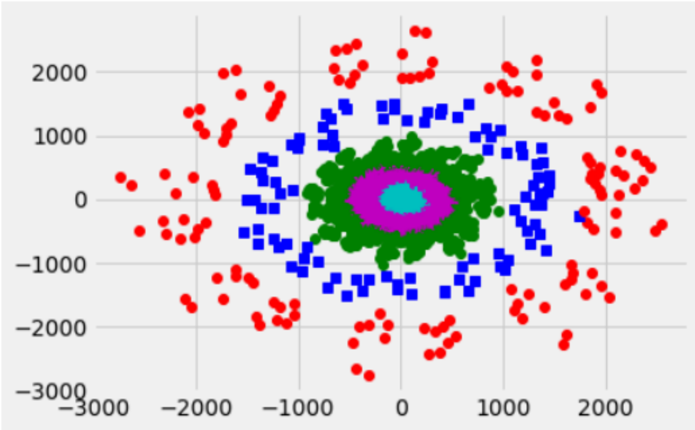


Fig. 3 Five types of passenger flow based on SOM classification

Then, based on the characteristics of working days and non working days, a correlation analysis was conducted for each day of the week, and the results are shown in **Fig. 4.**

Fig. 4 Correlation analysis for each day of the week

3.2. Construction of Combination Prediction Model Examples

The PSO-SVM and LSTM models used to predict passenger flow at Yuantong Station incorporate input features such as historical passenger data, weekday indicators, and weather factors. The composite forecasting model includes predictions from PSO-SVM and LSTM, weekday indicators, weather factors, and passenger flow classification values. Different sets of input features are employed to construct these models.

For the PSO-SVM prediction model, specific input features are extracted as shown in **Table 1** after processing. In this context, the target prediction period is defined as the t-th time segment of the d-th day, where $xd(t)$ represents the inbound or outbound passenger flow at that time, serving as the output variable $xd(t)$.

Table 1. Description of single model passenger flow input characteristics

Variable	Relevant characteristics
$xd{-}7(t)$	Passenger flow during the same period of the first week
$xd{-}14(t)$	Passenger flow during the same period in the first two weeks
$xd{-}21(t)$	Passenger flow during the same period in the first three weeks
$xd{-}28(t)$	Passenger flow during the same period in the first four weeks
$\mathrm{w}(d)$	Whether it is a workday characteristic
$t(d)$	Temperature
$r(d)$	Precipitation
$\mathrm{v}(d)$	Wind speed

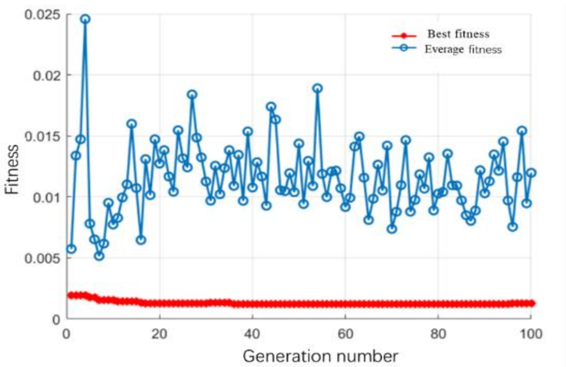
Based on the study of SVM models and existing research, the selection of the RBF kernel as the model's kernel function is confirmed due to its strong local learning ability and generalization capability. Following the algorithmic steps, parameter optimization of the core parameters is conducted using the PSO algorithm. Initially, the population is initialized, relevant parameters are defined, and initial particles and velocities are generated. Iterative optimization updates are performed to achieve individual and swarm best solutions. The loss function is set as Mean Squared Error (MSE), aiming to minimize MSE to identify the optimal parameter combination within the population for training and prediction of the PSO-SVM model.

Through iterative parameter tuning, specific experimental parameter settings are finalized as detailed in **Table 2**.

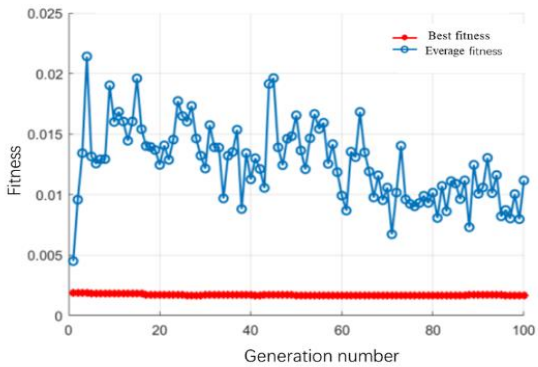
Table 2. PSO-SVM Core Parameter Values

parameter	value	parameter	value
c1	1.5	popcmax	1000
c2	1.7	popcmin	0.1
maxgen	100	popgmax	1000
sizepop	20	popgmin	0.001

Using c as the penalty factor and g as the kernel parameter, the optimal combinations of c and g for inbound passenger flow training were obtained through PSO global search, with 6.0762 and 3.1711, respectively, and an MSE of 0.0013. The optimal combinations of c and g for outbound passenger flow training were 11.7611 and 13.0241, respectively, with an MSE of 0.0017. Then pass the result down to train the SVM model for prediction. **Fig. 5** shows the population fitness curves for entering and exiting the station for optimization.



(a) Station entry optimization results



(b) Outbound optimization results

Fig. 5 Population fitness curve chart for station entry and exit optimization

For LSTM prediction models, their input features are the same as PSO-SVM. The main structure of the model consists of an input layer, a hidden layer, a fully connected layer, and an output layer.

This article confirms the setting of a single-layer hidden layer through comparative experiments., Then, based on the coefficient of determination *R* 2, the final confirmed number of entry model nodes is 80, and the number of exit model nodes is 90.

Table 3. Hidden layer node count performance

Station entry model		Outbound model	
Number of nodes	R^2	Number of nodes	R^2
60	0.9304	60	0.9412
70	0.9048	70	0.9383
80	0.9533	80	0.9409
90	0.9423	90	0.9483
100	0.9491	100	0.933

However, it may still fall into local optima. During training, stochastic gradient descent or momentum techniques are commonly employed. Considering the critical role of learning rate in model optimization, too small a learning rate may hinder convergence, while too large a rate can lead to fluctuating loss functions. Selecting an appropriate learning rate for gradient descent and its variants poses challenges. Momentum methods can strengthen relevant directions while weakening irrelevant ones, promoting stable convergence and reducing oscillations. The AdaGrad algorithm updates gradients adaptively based on the scale of sparse and frequent data, while the Adam algorithm combines these advantages by independently adapting learning rates using first and second moment estimates. [25-27]

Based on the passenger flow data in this section, the optimization algorithm adopts Adam, with Mean Squared Error (MSE) as the loss function and 500 iterations specified. **Table 4** details the critical parameters for LSTM, while **Fig. 6** and **Fig. 7** depict the iterative convergence processes for inbound and outbound traffic, respectively.

Table 4. The values of key parameters

Parameter	Value
MaxEpochs	500
GradientThreshold	1
InitialLearnRate	0.005
LearnRateDropFactor	0.2
Verbose	0
L2	0.004

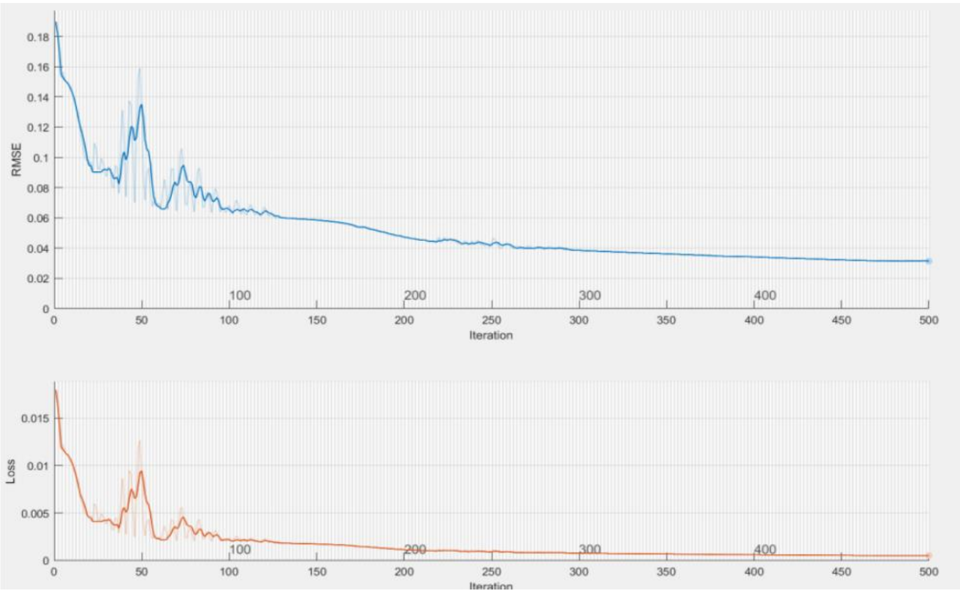


Fig. 6 The iterative convergence process of entering the station

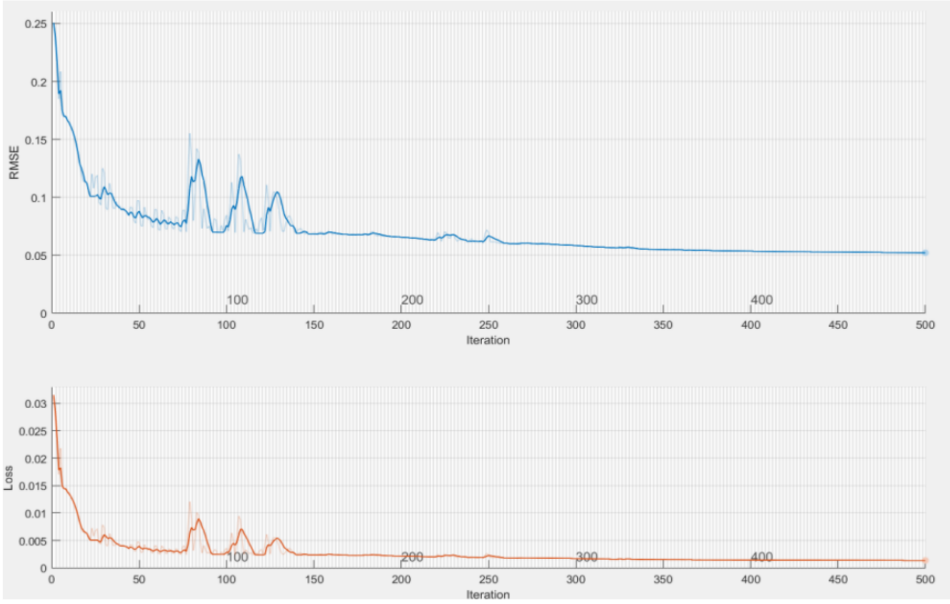


Fig. 7 The iterative convergence process of outbound

In the combination model section, the passenger flow dimension of its input features first includes the predicted values of PSO-SVM and LSTM, respectively. In addition, to ensure the accuracy of the prediction results, other influencing factor indicators were also inputted and assigned weights together for training, as shown in **Table 5**.

Table 5. Description of passenger flow input characteristics in combination model

Variable	Relevant Characteristics
$xSVM$	Prediction of passenger in PSO-SVM
$xLSTM$	Prediction of passenger in LSTM
$\mathrm{w}(d)$	Whether it is a workday characteristic
$o(d)$	Value of passenger flow classification
$t(d)$	Temperature
$r(d)$	Precipitation
$\mathrm{v}(d)$	Wind speed

The input layer is set to 7 nodes, and the output layer to 1 node, producing the combined predicted values. Backpropagation (BP) similarly relies on the processing capability of the hidden layer with increasing computational complexity and training speed as the number of layers increases. Through experimental comparison, it was found that a single hidden layer is sufficient to address the data issues in this study when an adequate number of neurons are employed in the hidden layer. Therefore, to minimize unnecessary computational overhead, a single hidden layer architecture was chosen. The number of hidden layer nodes was determined based on empirical formulas and previous research, confirmed by the coefficient of determination $R2$ of the final results. As shown in Table 6, it was determined that the number of nodes in the inbound model is 10 and in the outbound model is 11.

Table 6. The number of nodes and determination coefficients in the inbound and outbound model

Station entry model		Outbound model	
Number of nodes	$R2$	Number of nodes	$R2$
8	0.9326	8	0.9291
9	0.951	9	0.892
10	0.9617	10	0.9352
11	0.9583	11	0.9518
12	0.9464	12	0.8897
13	0.9341	13	0.8531

After continuous experimental adjustments, Table 7 shows the values of key parameters for the BP combination model, and **Fig. 8** shows the fitting results for both inbound and outbound training.

Table 7. The values of key parameters in the BP combination model

parameter	value	parameter	value
TF1	logsig	mc	0.3
TF2	purelin	epoch	300
lr	0.1	goal	1e-5

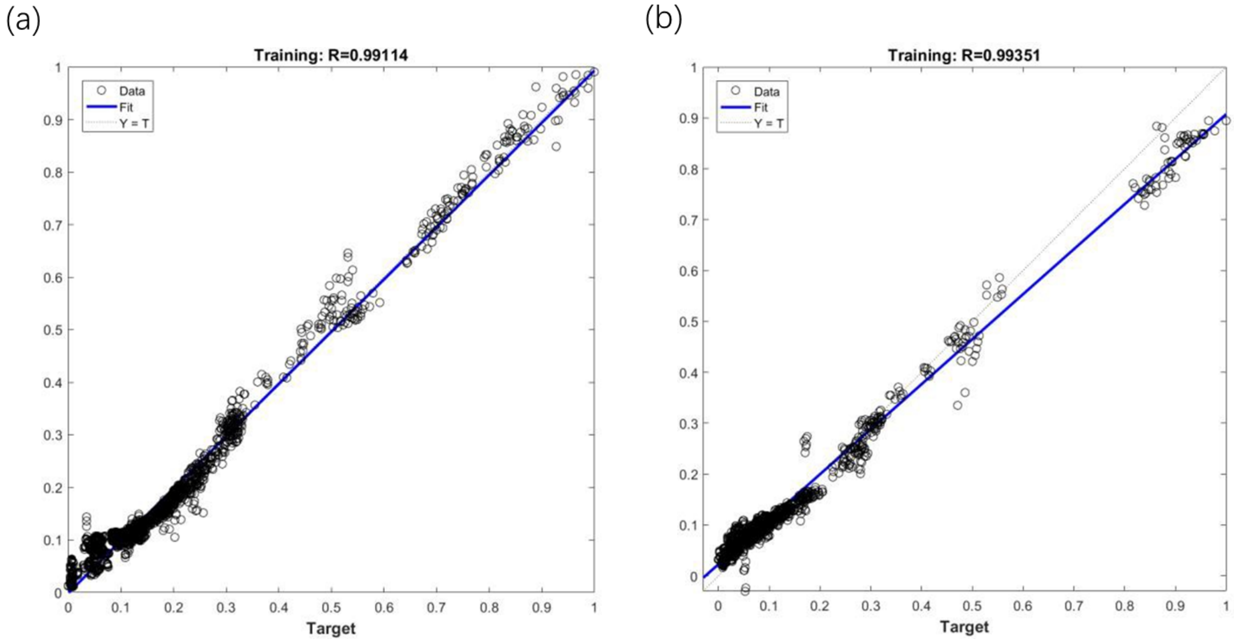


Fig. 8 Fitting situation of entrance and exit training

3.3. Analysis of Experimental Results

Fig. 9 shows a comparison between the predicted passenger flow of three models and the actual situation. And the comparison **table 8** of passenger flow evaluation indicators. The error after visible combination is basically lower than that of PSO-SVM and LSTM models.

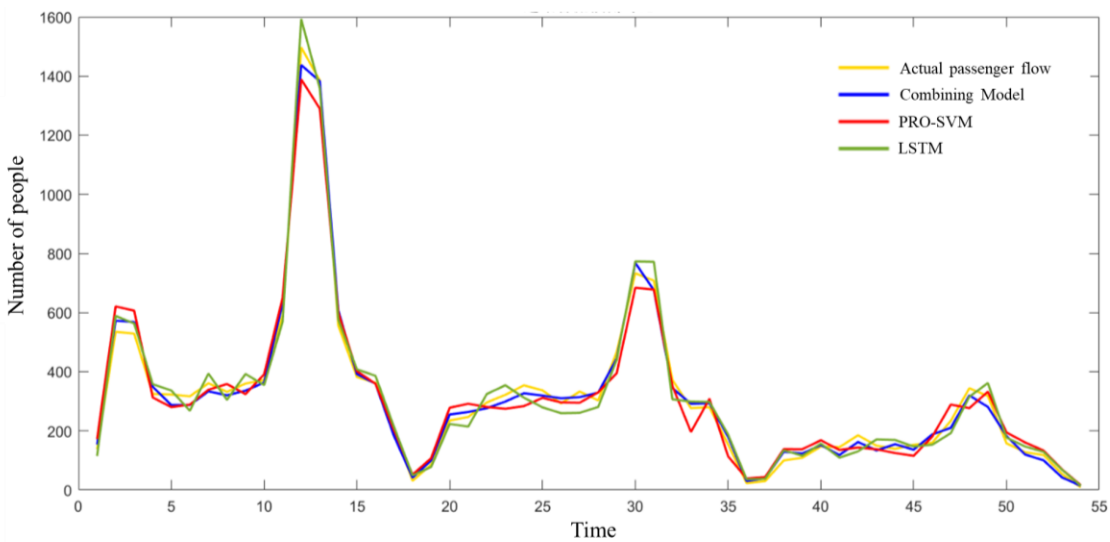


Fig. 9 Comparison between the prediction of inbound passenger flow by various models and the actual situation

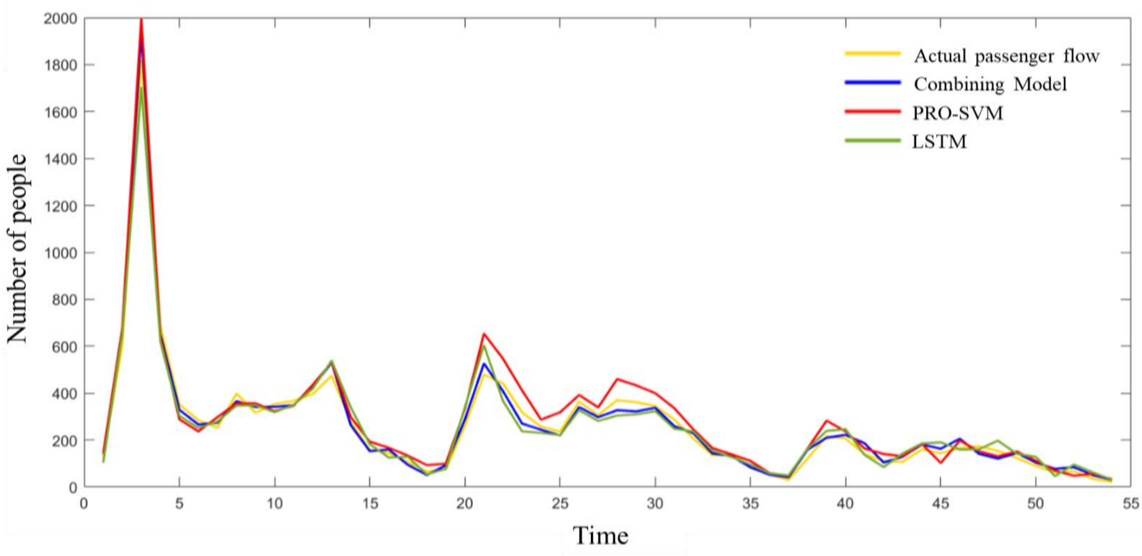


Fig. 10 Comparison between the prediction of outbound passenger flow by various models and the actual situation

Comparing the prediction errors of the three models in different time periods, it can be seen that the combined error is basically lower than that of the PSO-SVM and LSTM models.

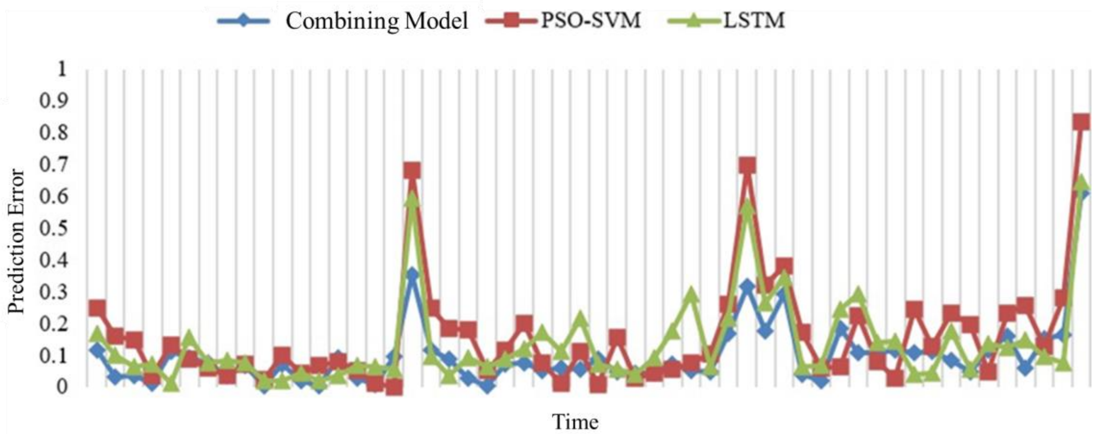


Fig. 11 Comparison of prediction errors of different models for inbound passenger flow in different time periods

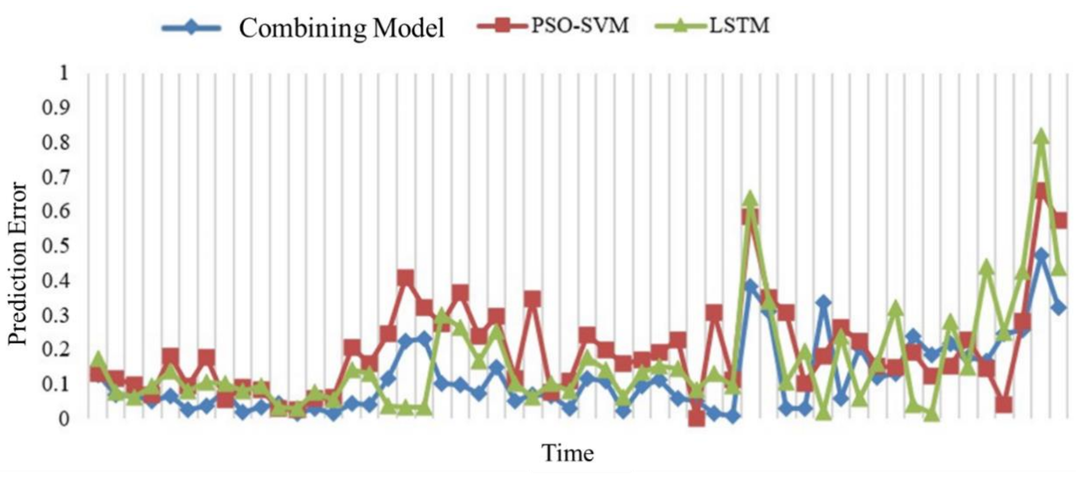


Fig. 12 Comparison of prediction errors of outbound passenger flow in different models at different time periods

It can also be seen from the overall evaluation index error statistics **table 9** that the results of the combined prediction model are better than those of the single model PSO-SVM and LSTM, with strong applicability. The combined performance effectively improves the prediction accuracy. The LSTM model outperforms the PSO-SVM model in overall performance, possibly due to its "memory unit" function. In addition, it was found in the prediction results of all models that the accuracy of the last few time periods predicted daily was relatively low, due to the relatively small passenger flow base of Yuantong Station in the last few time periods at night.

Table 8. Comparison of evaluation indicators for inbound passenger flow among various models

Models	MAE	MAPE	RMSE
Combining Model	21.33	10.87%	**26.29**
PSO-SVM	30.64	14.68%	**38.03**
LSTM	**27.58**	**12.71%**	**33.41**

Table 9. Comparison of evaluation indicators for outbound passenger flow among various models

Models	MAE	MAPE	RMSE
Combining Model	28.17	12.94%	34.10
PSO-SVM	38.98	17.79%	49.45
LSTM	31.68	15.59%	40.99

4. Conclusion

This paper proposes a passenger flow data combination preprocessing method combining self-organizing SOM algorithm and isolated forest method based on the passenger flow characteristics of Nanjing Metro Line 10, and constructs passenger flow input features. SVM and LSTM were used, and the particle swarm optimization (PSO) algorithm was applied to optimize the core parameter searching ability of SVM. PSO-SVM and LSTM short-term passenger flow prediction models were constructed separately. A short-term passenger flow combination prediction model based on PSO-SVM and LSTM was proposed, and the BP neural network algorithm was used for weight adaptation to improve the prediction accuracy. The experimental results show that the combination method has better predictive indicators than other methods. This combination model provides a new and effective approach for the further development of passenger flow forecasting methods.

Reference

[1] Han Linfei, Zhang Shenghai, Gao Meng Review and Reflection: The Impact of Soviet Urban Planning on Beijing Urban Planning in the 1950s Beijing Planning and Construction, 2009 (5): 15-20.

[2] Zhong Jiru Research on the Impact of Urban Built Environment on the Passenger Flow Characteristics of Rail Transit [D] Dalian University of Technology, 2019.

[3] Singhal A, Kamga C, Yazici A. Impact of weather on urban transit ridership[J]. Transportation research part A: policy and practice, 2014,69:379-391.

[4] Ding Shichao Research on the Impact of Weather Factors on High Speed Rail Passenger Flow [D] Beijing Jiaotong University, 2019.

[5] Dong H, Ye J. Passenger Flow Analysis Based on AFC Data of Shanghai Urban Rail Transit in People's Square Station[M]//ICTE 2013: Safety, Speediness, Intelligence, Low-Carbon, Innovation. 2013:736-744.

[6] Kitamura R, Fujii S, Pas E I. Time-use data, analysis and modeling: toward the next generation of transportation planning methodologies[J]. Transport Policy, 1997,4(4):225-235.

[7] Litman T. Generated traffic and induced travel[M]. Victoria Transport Policy Institute Victoria, BC, Canada, 2017.

[8] Jian J, De-ming Y. Analysis of Bus Diversity Fare Vote System and Passenger Flow Effect[J]. Journal of Transportation Systems Engineering and Information Technology, 2010,10(3):115.

[9] Zhang Y, Xie Y. Forecasting of short-term freeway volume with v-support vector machines[J]. Transportation Research Record, 2007,2024(1):92-99.

[10] Jianjun M, Pengfang C, Decang L I, et al. Summary of Research on Railway Freight Volume Forecast.[J]. Railway Standard Design, 2022,66(10).

[11] Guo Wen Research on Short term Passenger Flow Prediction Method for Rail Transit Based on Support Vector Machine [D] Suzhou University, 2019.

[12] Zhu Bingjie Application of Improved K-Nearest Neighbor Algorithm in Urban Rail Transit Passenger Flow Prediction [D] Beijing Jiaotong University, 2019.

[13] Cha Min Research on Public Transport Passenger Flow Prediction Method Based on Spark Platform [D] University of Electronic Science and Technology of China, 2017.

[14] Kennedy J, Eberhart R. Particle swarm optimization: Proceedings of ICNN'95-international conference on neural networks[C], 1995. ieee.

[15] Ahmed M S, Cook A R. Analysis of freeway traffic time-series data by using Box-Jenkins techniques[M]. 1979.

[16] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks: 2013 IEEE international conference on acoustics, speech and signal processing[C], 2013. Ieee.

[17] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997,9(8):1735-1780.

[18] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification: Proceedings of the IEEE international conference on computer vision [C], 2015.

[19] Zhang Y, Xie Y. Forecasting of short-term freeway volume with v-support vector machines [J]. Transportation Research Record, 2007,2024(1):92-99.

[20] Zhu J Z, Cao J X, Zhu Y. Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections [J]. Transportation Research Part C: Emerging Technologies, 2014,47:139-154.

[21] Jian J, De-ming Y. Analysis of Bus Diversity Fare Vote System and Passenger Flow Effect[J]. Journal of Transportation Systems Engineering and Information Technology, 2010,10(3):115.

[22] Kalkstein A J, Kuby M, Gerrity D, et al. An analysis of air mass effects on rail ridership in three US cities[J]. Journal of transport geography, 2009,17(3):198-207.

[23] Guo Z, Wilson N H, Rahbee A. Impact of weather on transit ridership in Chicago, Illinois[J]. Transportation Research Record, 2007,2034(1):3-10.

[24] Stover V W, McCormack E D. The impact of weather on bus ridership in Pierce County, Washington[J]. Journal of Public Transportation, 2012,15(1):95-110.