Journal of Southeast University (English Edition) Vol. 38, No. 4, pp. 408 - 417 Dec. 2022 ISSN 1003—7985

Probabilistic interval prediction of metro-to-bus transfer passenger flow in the trip chain

Shen Jin[1] Zhao Jiandong[1,2] Gao Yuan[3] Feng Yingzi[1] Jia Bin[1]

([1]School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China)
([2]Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China)
([3]School of Traffic and Transportation, Northeast Forestry University, Harbin 150040, China)

Abstract: To accurately analyze the fluctuation range of time-varying differences in metro-to-bus transfer passenger flows, the application of a probabilistic interval prediction model is proposed to predict transfer passenger flows. First, bus and metro data are processed and matched by association to construct the basis for public transport trip chain extraction. Second, a reasonable matching threshold method to discriminate the transfer relationship is used to extract the public transport trip chain, and the basic characteristics of the trip based on the trip chain are analyzed to obtain the metro-to-bus transfer passenger flow. Third, to address the problem of low accuracy of point prediction, the DeepAR model is proposed to conduct interval prediction, where the input is the interchange passenger flow, the output is the predicted median and interval of passenger flow, and the prediction scenarios are weekday, non-workday, and weekday morning and evening peaks. Fourth, to reduce the prediction error, a combined particle swarm optimization (PSO)-DeepAR model is constructed using the PSO to optimize the DeepAR model. Finally, data from the Beijing Xizhimen subway station are used for validation, and results show that the PSO-DeepAR model has high prediction accuracy, with a 90% confidence interval coverage of up to 93.6%.

Key words: urban traffic; probabilistic interval prediction; deep learning; metro-to-bus transfer passenger flow; trip chain
DOI: 10.3969/j.issn.1003-7985.2022.04.010

As urban residents' multimodal trips are developing rapidly, fully mining passenger travel data, extracting travel chain features, and conducting interchange passenger flow prediction are the current research hotspots of multimodal trips, which are of reference significance for improving residents' trip efficiency and coping with holiday peak passenger flow induction and emergency passenger flow evacuation.

In their study of trip chains, Adler et al.[1] investigated passenger trip chains from the perspective of utility maximization, Kondo et al.[2-3] analyzed several factors affecting trip chains, Qi et al.[4] explored the relationship between transport modes and trip chains using nested logit models, and Wang et al.[5] proposed a user-balanced transport allocation model based on trip chains. In terms of public transport interchange passenger flow, Wang et al.[6] established a simulation model of transfer organization, Wang et al.[7] proposed a coordinated design of operating routes between multiple transfer points to reduce transfer costs, Zhang[8] analyzed the coordination of metro-to-bus transfers, and Xiong et al.[9] predicted the transfer passenger flow based on Kalman filtering. In terms of passenger flow prediction, point prediction methods for short-term passenger flow mainly include autoregressive integrated moving average (ARIMA)[10], convolutional neural network[11], support vector machine (SVM)[12], and long short-term memory[13] models.

In contrast to outputting specific point forecasts, probabilistic interval forecasts output a range of possible fluctuations in passenger flow, and the reliability of the forecast results is measured by a confidence metric. The scheduling planner selects the prediction result under the best confidence level according to the need to complete the safe and reliable scheduling of public transport; thus, the probabilistic interval prediction for predicting the probability distribution of passenger flow has attracted considerable attention from scholars. Zhang[14] investigated short-term traffic flow interval prediction based on the gray system theory. Zhu et al.[15] determined the traffic flow probability intervals based on the Bayesian network posterior distribution. Tong et al.[16] predicted the trend and range of traffic flow change by optimizing the SVM model. Thus, the probabilistic interval prediction is closer to the essence, and the accuracy of the prediction is higher.

However, only a few studies of public transport trip chains, transfer passenger flow, and time series probabilistic interval prediction have been conducted, and the research on probabilistic interval prediction of metro-to-bus

Received 2021-10-30, **Revised** 2022-06-20.

Biographies: Shen Jin (1997—), female, Ph. D. graduate; Zhao Jiandong (corresponding author), male, doctor, professor, zhaojd@bjtu.edu.cn.

Foundation items: The National Key Research and Development Program of China (No. 2019YFB160-0200), the National Natural Science Foundation of China (No. 71871011, 71890972/71890970).

Citation: Shen Jin, Zhao Jiandong, Gao Yuan, et al. Probabilistic interval prediction of metro-to-bus transfer passenger flow in the trip chain [J]. Journal of Southeast University (English Edition), 2022, 38(4): 408 - 417. DOI: 10.3969/j.issn.1003-7985.2022.04.010.

transfer passenger flow has yet to be enriched.

DeepAR is a time series prediction method based on deep learning proposed by Salinas et al. [17], which outputs the prediction results in the form of probability distributions, provides interval estimates, and achieves more accurate time series prediction. Yan et al. [18] evaluated the uncertainty of Web traffic prediction based on the Prophet-DeepAR model. Zhu et al. [19] verified the effectiveness of the DeepAR model in predicting the probability distribution of time series. Li et al. [20] used the DeepAR model to predict the remaining life of an aeroengine. The DeepAR prediction model, which has a wide range of applications but has not been extensively employed in passenger flow prediction, is mainly utilized for time series prediction.

Therefore, in this study, a combined particle swarm optimization (PSO)-DeepAR[21] model is constructed to predict the probability interval of metro-to-bus transfer passenger flow based on multisource public transport data, extract the public transport trip chain, and determine the metro-to-bus transfer passenger flow. Moreover, the actual data of the Beijing Xizhimen subway station are used as an example to compare and verify the accuracy of the combined model.

1 Public Transportation Swipe Card Data Preprocessing

In this study, we collected bus and metro swipe card data, bus and metro line data, and bus and metro station data of Beijing in April and May 2018. Among them, the bus weekday daily swipe card data are approximately 8×10^6, and the metro weekday daily swipe card data are approximately 5×10^6. The names of the data fields used are shown in Tab. 1.

Tab. 1 Names of the swipe card data fields

Serial number	Field name	Meaning
1	GRANT_CARD_CODE	Card number
2	ON_TIME	Boarding time
3	OFF_TIME	Drop-off time
4	ON_LINE	Boardingline
5	OFF_LINE	Drop-off line
6	ON_STATION	Boarding station
7	OFF_STATION	Drop-off station
8	CARD_TYPE	Card type

These data originate from different data systems with different field name identifications. First, bus and metro swipe card data are preprocessed. Bus swipe card data are cleaned to remove the data with the same in and out stations, earlier out time than in time, within nonoperating time, long trip time, not swiped card when getting off, and line stations that cannot be matched. The rail card data are cleaned to remove the data with the same up and down stations, earlier time of getting off than time of getting on, nonoperating time, and long trip time. After cleaning, the valid data of the bus accounted for 88%, and the valid data of the metro accounted for 97%. The cleaned data are matched to obtain the card data integration table.

2 Trip Chain Identification and Transfer Passenger Flow Analysis

2.1 Trip chain structure extraction

A trip phase[22] is a process in which a tripper uses the same mode of transportation to get from the origin to the destination. A public transportation trip chain[23] is a complete trip process consisting of one or more public transportation trip stages in the order of occurrence from the trip origin station to the destination station. That is, a complete public transportation trip can be composed of only one trip stage, or several trip stages can be separated by transfer points. The final constituted public transportation trip chain can reproduce the specific process of each public transportation trip of the tripper. The trip chain and trip phase relationships are shown in Fig. 1.

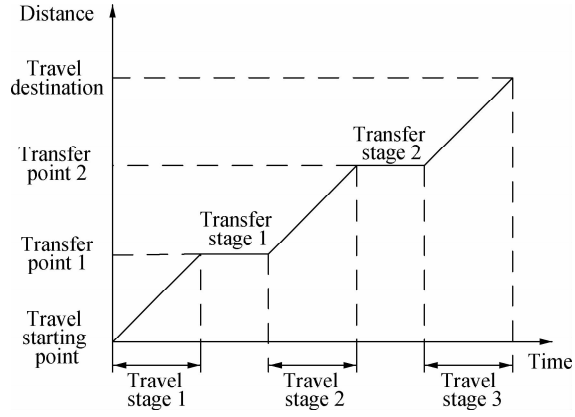

Fig. 1 Schematic diagram of the trip chain

The extraction of the public transportation trip chain structure is the basis for transfer point determination and transfer passenger flow extraction. In this study, we use the transaction time difference between two adjacent swipe records of passengers with the same card number in the public transportation swipe data integration table to determine whether the two trip stages are transfer relationships and whether each swipe record, i. e., each trip stage, belongs to the same trip chain. Then, the structure of each trip chain is extracted.

Transfers in the public transportation trip chain of passengers are divided into three types, namely, bus to bus, bus to metro, and metro to bus. In this study, the bus trips of passengers are denoted as B, and the metro trips are denoted as M.

The transfer time of metro to bus includes the time passengers spend walking to the bus station after exiting the metro station and the time spent waiting at the bus stop. The former can be determined based on the average dis-

tance of bus stations around the metro station, and the latter can be determined based on the average waiting time model for metro-to-bus transfer[24]. In general, a bus station can be reached within the city by a 5-min walk from the metro station[25]; the shortest bus departure interval during peak hours is generally less than 10 min, taking into account that passengers will exhibit other behaviors, such as simple shopping during the transfer process; and the maximum transfer time for metro-to-bus transfer is set at 30 min. Analysis of the distribution of the card time difference between the two trip stages in 30 min showed that the cumulative frequency of the card time difference is 95% of the 23-min metro-to-bus transfer time threshold, i. e., when the same trip chain of the latter trip stage bus trip boarding time minus the previous trip stage metro trip off time is less than 23 min, the passenger is determined to have metro-to-bus behavior. The schematic diagram of the calculation method is shown in Fig. 2. Similarly, we obtained the bus-to-bus time threshold of 26 min and the bus-to-metro time threshold of 18 min using the calculation method.

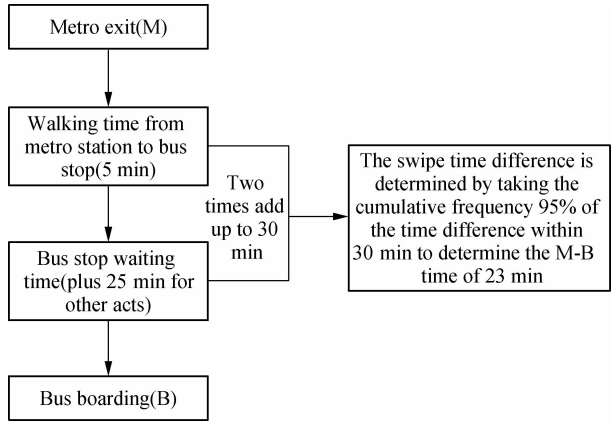

Fig. 2 Time threshold identification diagram of metro-to-bus transfer

Based on the three time thresholds, we extract the trip chain structure of passengers and add two fields, i. e., "CHAIN" and "STAGE," to the public transportation card data integration table. "CHAIN" denotes the *i*-th trip chain of the passenger, and "STAGE" denotes the *j*-th trip stage of the passenger. After tagging, we obtain the public transportation trip chain information tags (see Tab. 2).

Tab. 2 Public transport trip chain structure marking

GRANT_CARD_CODE (card number)	ON_TIME (boarding time)	OFF_TIME (drop-off time)	…	MODE (trip pattern)	CHAIN (trip chain)	STAGE (trip stage)
1123	2018-04-07T09:24	2018-04-07T09:46		B	1	1
1123	2018-04-07T20:04	2018-04-07T20:22	…	B	2	1
1149	2018-04-07T06:22	2018-04-07T06:42	…	B	1	1
1149	2018-04-07T06:44	2018-04-07T07:38	…	M	1	2
1149	2018-04-07T17:43	2018-04-07T08:33	…	M	2	1
1149	2018-04-07T18:43	2018-04-07T19:03	…	B	2	2

Taking the card swipe data on April 7, 2018, as an example for statistical analysis, a total of 3.82×10^6 trips were made on that day, with a total of 9.58×10^6 card swipes, an average of 2.39 swipes per person, and a total of 6.61×10^6 trip chains, of which 95% of people's trip chains contained less than or equal to 4 trip stages. That is, most people generally do not make more than three transfer times in one trip. The number of trip structure type statistics of public transportation trip chains of trippers is shown in Tab. 3.

Tab. 3 Statistics of the number and proportion of different types of trip structures on April 7, 2018

Trip structure	Number	Percentage/%
M	4 667 704	46.35
B	3 594 147	35.69
B-B	805 332	8.00
B-M	408 929	4.06
M-B	4 036 124	4.01
B-B-B	98 275	0.98
B-M-B	42 588	0.42
B-B-B-B	14 625	0.15
M-B-B	12 227	0.12
B-B-M	11 345	0.11
M-B-M	1 737	0.02
B-M-B-B	1 627	0.02
B-B-M-B	1 454	0.01
M-B-B-B	920	0.01
B-B-B-M	832	0.01
M-B-B-M	760	0.01
M-B-M-B	189	0
B-M-B-M	182	0

2.2 Extraction and analysis of metro-to-bus transfer passenger flow

Beijing's public transportation network covers a large and detailed area, and frequent transfers are very rare. In general, passengers do not transfer many times, and considering that passengers who transfer more than three times may be engaged in special jobs or have problems with the public transportation system statistics, we extracted the card records of passengers with trip stages less than or equal to four for transfer passenger flow statistics. To clarify the research object, the Xizhimen metro station

was selected to obtain the metro-to-bus transfer passenger flow statistics. For example, on April 9, 2018, 58 691 people got on and 58 759 people got off at Xizhimen metro station; 7 229 of the people that exited from the Xizhimen metro station changed to buses; given that the metro generally departs every 5 min, the passenger flow was aggregated at a time granularity of 5 min[26]. The metro-to-bus passenger flow statistics for one week from April 9, 2018, to April 15, 2018, are shown in Fig. 3.

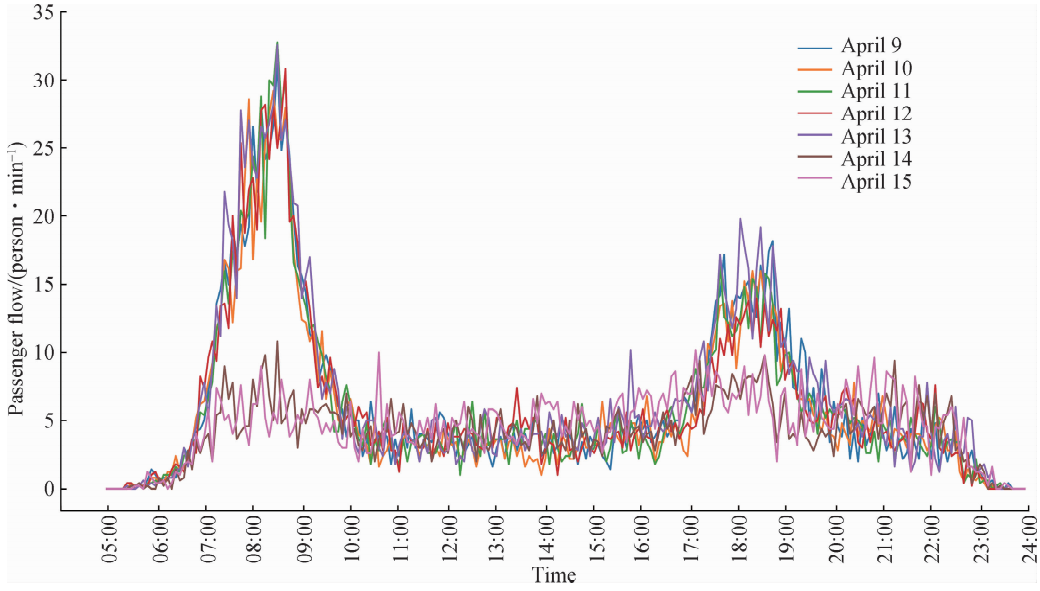

Fig. 3 Display of Xizhimen metro station connection and transfer volume

Fig. 3 shows the weekday transfer passenger flow from April 9, 2018, to April 13, 2018, and the nonworking day transfer passenger flow from April 14, 2018, to April 15, 2018. The weekday transfer passenger flow exhibits an obvious morning and evening peak regularity, whereas the nonworking day passenger flow shows an inconspicuous morning and evening peak passenger flow regularity. Thus, the prediction scenarios are categorized into weekdays and nonworking days. To improve the prediction accuracy, decision-makers need to reasonably plan and induce passenger flow during the peak and low peak periods, deploy station security measures in advance, and improve transfer efficiency. This study separates the weekday morning and evening peak transfer passenger flows for separate predictions.

3 Probabilistic Interval Prediction Model for Metro-to-Bus Transfer Passenger Flow

3.1 DeepAR probabilistic interval forecasting model

DeepAR is an autoregressive recurrent neural network (RNN) time series model, which is an RNN model with hidden states. The model predicts the probability distribution of $\tilde{z}_t$ based on the autoregressive RNN. $\tilde{z}_t$ denotes the predicted values of the model at the time step and t_0 denotes the prediction start moment. The training starts with the input layer for data $\boldsymbol{X}_t$ input, and the metro-to-bus transfer passenger flows are inputted into the model. At each time step t, the input to the network includes the value $\boldsymbol{X}_{t-1}$ taken at the previous time step and the state $\boldsymbol{h}_{t-1}$ at the previous time step.

The network layer is a neural network containing hidden states, where $\boldsymbol{z}_t$ is the small batch input of time step t in the sequence, $\boldsymbol{h}_t$ is the hidden variable of that time step, $\boldsymbol{h}_{t-1}$ is the hidden variable of the previous time step, and $\boldsymbol{h}_t = \boldsymbol{h}(\boldsymbol{h}_{t-1}, \boldsymbol{Z}_{t-1})$ is the current state, which is computed in the network layer. A new weight parameter W_{hh} is introduced to determine how the current time step uses the hidden variable of the previous time step. The hidden variable of time step t of the hidden variable $\boldsymbol{h}_t$ is jointly determined using the input of the current time step and the hidden variable of the previous time step. The hidden variable can be used to determine the state or network memory of the current time step; thus, the hidden variable is also called the hidden state.

$$\boldsymbol{h}_t = \phi(\boldsymbol{Z}_t W_{\mathrm{xh}} + \boldsymbol{h}_{t-1} W_{\mathrm{hh}} + \boldsymbol{b}_{\mathrm{h}}) \tag{1}$$

$$\boldsymbol{O}_t = \boldsymbol{h}_t W_{\mathrm{hq}} + \boldsymbol{b}_{\mathrm{q}} \tag{2}$$

The hidden state at the current time step is defined using the hidden state from the previous time step, and Eq. (1) is computed cyclically. The output layer is calculated using Eq. (2).

The computational logic of the network layer RNN at three adjacent time steps is shown in Fig. 4. The computation of the hidden state can be viewed as a fully connected layer with the activation function ϕ after linking the input $\boldsymbol{Z}_t$ with the hidden state $\boldsymbol{h}_{t-1}$ of the previous time step. The output of this connected layer is the hidden state $\boldsymbol{h}_t$ of the current time step, and the model parameters are the link between W_{xh} and W_{hh} with a deviation of b_{h}.

The hidden state $\boldsymbol{h}_t$ of the current time step t will be involved in the calculation of the hidden state $\boldsymbol{h}_{t+1}$ of the next time step $t+1$ and inputted into the fully connected transport layer of the current time step, where W_{xh}, W_{hh}, and W_{hq} are the weight parameters and b_{h} and b_{q} are the deviations.

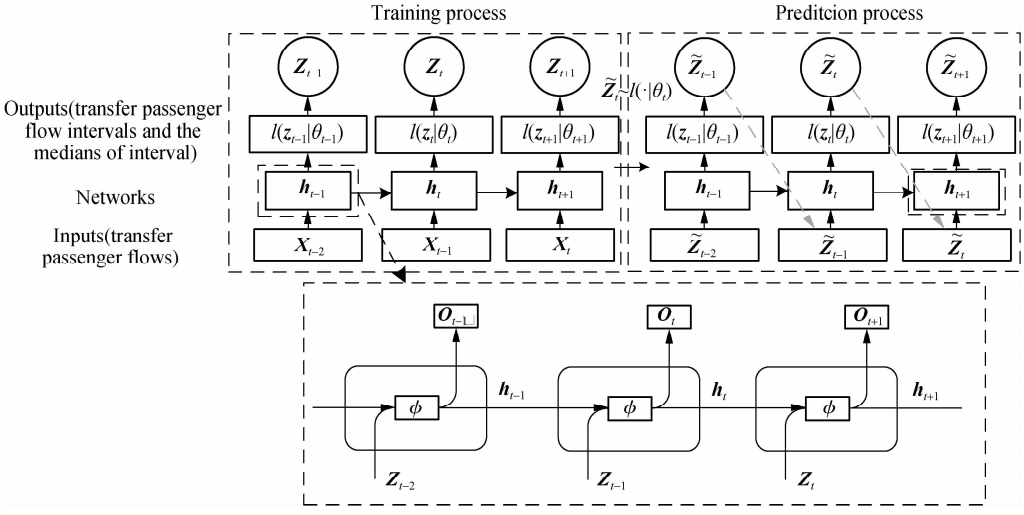

Fig. 4 DeepAR model training and prediction models

After obtaining the output value $\boldsymbol{O}_t$, the parameter θ_t of the likelihood function $l(z_t \mid \theta(\boldsymbol{O}_t))$ is calculated, and the parameters of the network are learned by maximizing the log likelihood. The final output value of the training is $\tilde{\boldsymbol{Z}}_t$, including the predicted transfer passenger flow intervals and the medians of intervals.

$$L(\theta) = \sum_t \lg l(z_t \mid \theta(\boldsymbol{O}_t)) \tag{3}$$

The detailed training process of the model is shown in Fig. 4, with the training process on the left and the prediction process on the right.

After the training is completed, the $t < t_0$ historical data is inputted into the network to obtain the initial state $\boldsymbol{h}_{t_0-1}$. Then, ancestor sampling can be used to obtain the prediction results: For $t_0, t_0+1, \ldots, T$, each time step is randomly sampled to obtain $\tilde{\boldsymbol{Z}}_t \sim l(\cdot \mid \theta_t)$, and this sampled value is used as the input for the next time step. Repeating this process yields a series of $t_0 \sim T$ values, which can be used to compute the desired target values, such as quantile and expectation.

The specific form of $\theta(\boldsymbol{O}_t)$ depends on the likelihood function $l(\boldsymbol{Z}_t \mid \theta_t)$, and the likelihood function needs to be selected based on the statistical characteristics of the data itself. The likelihood function that best matches the statistical properties of the data needs to be selected, and the commonly used likelihood functions are Gaussian distributions for real-valued data. If a Gaussian distribution is selected, then the Gaussian distribution is parameterized using the mathematical expectation μ and deviation σ, i.e., $\theta = (\mu, \sigma)$, where the mathematical expectation μ is derived using the affine transformation function of the network output, and the deviation σ is obtained using the affine transformation of the following activation functions to ensure that the variance is greater than 0, where

$$l_\sigma(z \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(\frac{-(z-\mu)^2}{2\sigma^2}\right) \tag{4}$$

$$\mu(\boldsymbol{O}_t) = w_u \boldsymbol{O}_t + \boldsymbol{b}_u \tag{5}$$

$$\sigma(\boldsymbol{O}_t) = \log(1 + \exp(w_\sigma \boldsymbol{O}_t + \boldsymbol{b}_\sigma)) \tag{6}$$

where w_u and w_σ are weight parameters; $\boldsymbol{b}_u$ and $\boldsymbol{b}_\sigma$ are the bias variables; and $\mu(\boldsymbol{O}_t)$ and $\sigma(\boldsymbol{O}_t)$ are the mean and standard deviation of the Gaussian distribution function, respectively.

The affine transformation function is used to compute the output of the fully connected layer, the result of which is the input final activation function. In this study, a Gaussian distribution is used as the likelihood function model. That is, instead of a common RNN model that predicts points directly, a probability is predicted, from which the predicted value is then obtained, and the range in which the point may occur is described using features of the probability distribution. In contrast to point prediction, which results in a single point or a specific value, probabilistic prediction is equivalent to predicting the probability distribution of that point and being able to use the features of the probability distribution to describe the range in which that point may occur.

3.2 PSO algorithm

The PSO algorithm is derived from the movements of a flock of birds searching for food to obtain an optimized solution. The basic idea is to identify the best position of each individual by analyzing its adaptability to the environment and move to the best position in the region using mutual collaboration and information sharing among each individual in the group. The algorithm treats each indi-

vidual in the group as a particle without volume and mass, and each particle flies in the given search space at a certain speed, which is dynamically iterated and updated according to the particle itself and the flight of its surrounding companions. In the process of solving the optimization problem, the particles dynamically adjust their velocities and positions to obtain the optimization solution. The process can be expressed as follows:

$$V_{ts}^{k+1} = \delta V_{ts}^{k} + e_1 R_1 (P_{ts1}^{k} - X_{ts}^{k}) + e_2 R_2 (P_{ts2}^{k} - X_{ts}^{k}) \quad (7)$$

$$X_{ts}^{k+1} = X_{ts}^{k} + V_{ts}^{k+1} \quad (8)$$

where V_{ts}^{k} is the velocity of the t-th particle in search space dimension s and iteration number k; δ is the inertia weight; X_{ts}^{k} is the position of the t-th particle in search space dimension s and iteration number k; P_{ts1}^{k} and P_{ts2}^{k} are the individual and global extremes of the t-th particle in search space dimension s and iteration number k, respectively; e_1 and e_2 are the acceleration factors, both of which are nonnegative constants; and R_1 and R_2 are constants in the $(0,1)$ interval.

3.3 PSO-DeepAR probabilistic interval prediction model

The modelling process for the PSO-DeepAR model is shown below.

① The DeepAR network structure is designed.

② The DeepAR network is trained.

③ The initial parameters of the optimal prediction model are determined.

④ The particle optimal position and individual are updated according to the size of the particle fitness value.

⑤ If the optimality search condition is satisfied, then the iteration is terminated; otherwise, the particle position and velocity are continuously updated until the termination condition is satisfied.

⑥ The optimal solution to the DeepAR model is assigned.

4 Case Analysis

4.1 Interval prediction performance indicators

We assume that the predicted value is $\hat{Y}_t = \{\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n\}$ and the true value is $\hat{Y}_t = \{\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n\}$. The predicted value is the median of the interval obtained by predicting the passenger flow. The following error evaluation metrics were used to measure the model interval prediction performance.

1) Mean absolute square error (MASE)

MASE reflects the superiority of probabilistic interval prediction over plain average prediction.

$$\mathrm{MASE} = \frac{1}{h} \frac{\sum_{t=1}^{h} |Y_t - \hat{Y}_t|}{\frac{1}{n-m}\sum_{t=m+1}^{n} |Y_t - Y_{t-m}|} \quad (9)$$

where h is the length of the prediction and m is the frequency of the time series.

2) Symmetric mean absolute percentage error (sMAPE)

sMAPE determines the magnitude of the difference between predicted and true values. The smaller the value, the smaller the error.

$$\mathrm{sMAPE} = \frac{1}{h}\sum_{t=1}^{h} \frac{|Y_t - \hat{Y}_t|}{(|Y_t| + |\hat{Y}_t|)/2} \quad (10)$$

3) Weighted quantile loss (wQuantileLoss[τ])

The wQuantileLoss[τ] metric is used to measure the accuracy of the model at a specified distribution point called quantile. This metric helps capture the inherent bias in each quantile, and the selection of a higher quantile better captures the peak passenger flow. The wQuantileLoss[τ] is expressed as

$$\mathrm{wQuantileLoss}[\tau] = 2\frac{\sum_{t=1}^{h}[\tau \max(Y_t - \hat{Q}_t^{(\tau)}, 0) + (1-\tau)\max(\hat{Q}_t^{(\tau)} - Y_t, 0)]}{\sum_{t=1}^{h}|Y_t|} \quad (11)$$

where τ is the quartile, $\tau \in [0.1, 0.99]$; $\hat{Q}_t^{(\tau)}$ is the quantile of the model prediction. In this study, we select the 0.9 quantile to identify the prediction accuracy, which indicates that the true value is expected to be lower than the predicted value 90% of the time. Moreover, the smaller the value is, the smaller the error.

4) Standardized deviation (ND)

The smaller the standardized deviation, the smaller the error.

$$\mathrm{ND} = \frac{\sum_{t=1}^{h}|Y_t - \hat{Y}_t|}{\sum_{t=1}^{h}|Y_t|} \quad (12)$$

5) Mean scaled interval score (MSIS)

The smaller the MSIS value, the smaller the error.

$$\mathrm{MSIS} = \frac{\frac{1}{n-m}\sum_{t=m+1}^{n}|Y_t - Y_{t-m}| + \frac{2}{\beta}(\hat{L}_t - Y_t)I_{\{Y_t < \hat{L}_t\}} + \frac{2}{\beta}(Y_t - \hat{U}_t)I_{\{Y_t > \hat{U}_t\}}}{\frac{1}{n-m}\sum_{t=m+1}^{n}|Y_t - Y_{t-m}|} \quad (13)$$

where β is the level of significance ($\beta \in [0,1]$); $\hat{U}_t$ is the upper bound of the prediction interval; $\hat{L}_t$ is the lower bound of the prediction interval; $I_{\{Y_t<\hat{L}_t\}}$ is the relationship between Y_t and $\hat{L}_t$ when $Y_t < \hat{L}_t$, $I_{\{Y_t<\hat{L}_t\}} = 1$ (otherwise, $I_{\{Y_t<\hat{L}_t\}} = 0$); and $I_{\{Y_t>\hat{U}_t\}}$ is the relationship between Y_t and $\hat{U}_t$ when $Y_t > \hat{U}_t$, $I_{\{Y_t>\hat{U}_t\}} = 1$ (otherwise, $I_{\{Y_t>\hat{U}_t\}} = 0$). $\frac{1}{h}\sum_{t=1}^{h}(\hat{U}_t - \hat{L}_t)$ penalizes the interval between upper and lower bounds, $\frac{2}{\beta}(\hat{L}_t - Y_t) I_{\{Y_t<\hat{L}_t\}}$ penalizes cases where the true value is below the lower bound, and $\frac{2}{\beta}(Y_t - \hat{U}_t)$ $I_{\{Y_t>\hat{U}_t\}}$ penalizes cases where the true value is above the upper bound.

6) α confidence interval coverage (Coverage[α])

Coverage[α] is the proportion of h-step predictions in which the true value Y_t is less than or equal to the α quantile of the prediction $\hat{Y}_t$. If the prediction is more accurate, the proportion should be closer to α.

$$\text{Coverage}[\alpha] = \frac{1}{h}\sum_{t=1}^{h} I_{\{\hat{Y}_t \geqslant Y_t\}} \tag{14}$$

where $I_{\{\hat{Y}_t \geqslant Y_t\}}$ is the relationship between Y_t and $\hat{Y}_t$ (when $Y_t \leqslant \hat{Y}_t$, $I_{\{\hat{Y}_t \geqslant Y_t\}} = 1$; otherwise, $I_{\{\hat{Y}_t \geqslant Y_t\}} = 0$) and α is quantile $\alpha \in [0,1]$.

4.2 Forecast analysis of metro-to-bus transfer passenger flow

The transfer flows based on the analysis of trip chain characteristics are inputted into the PSO-DeepAR model. The final optimal parameters of the DeepAR model optimized using the PSO algorithm are a learning time step of 12, learning rate of 10^{-3}, number of hidden layers of 1, number of RNN fiducials per layer of 40, and batch size of 100. The model prediction results are obtained in 50 cycles. The model outputs for weekday, non-weekday, and weekday morning peak (7:30 to 9:30) and evening peak (17:00 to 19:00) transfer passenger flow intervals. The medians of intervals are shown in Fig. 5.

Taking Fig. 5 (a) as an example analysis, the blue line denotes the actual passenger flow, the green line denotes the median of the passenger flow forecast interval, the dark green interval denotes the 50% confidence prediction probability, and the light green interval denotes the 90% confidence prediction interval, i.e., the probability that 90% of the predicted passenger flow falls within the light green interval is 93.60%. Thus, the confidence probability is higher at 90%, and all passenger flows fall within the prediction interval range. The comparison of the 50% and 90% prediction accuracies shows that the 90% confidence probability prediction accuracy is higher than the 50% confidence probability prediction accuracy.

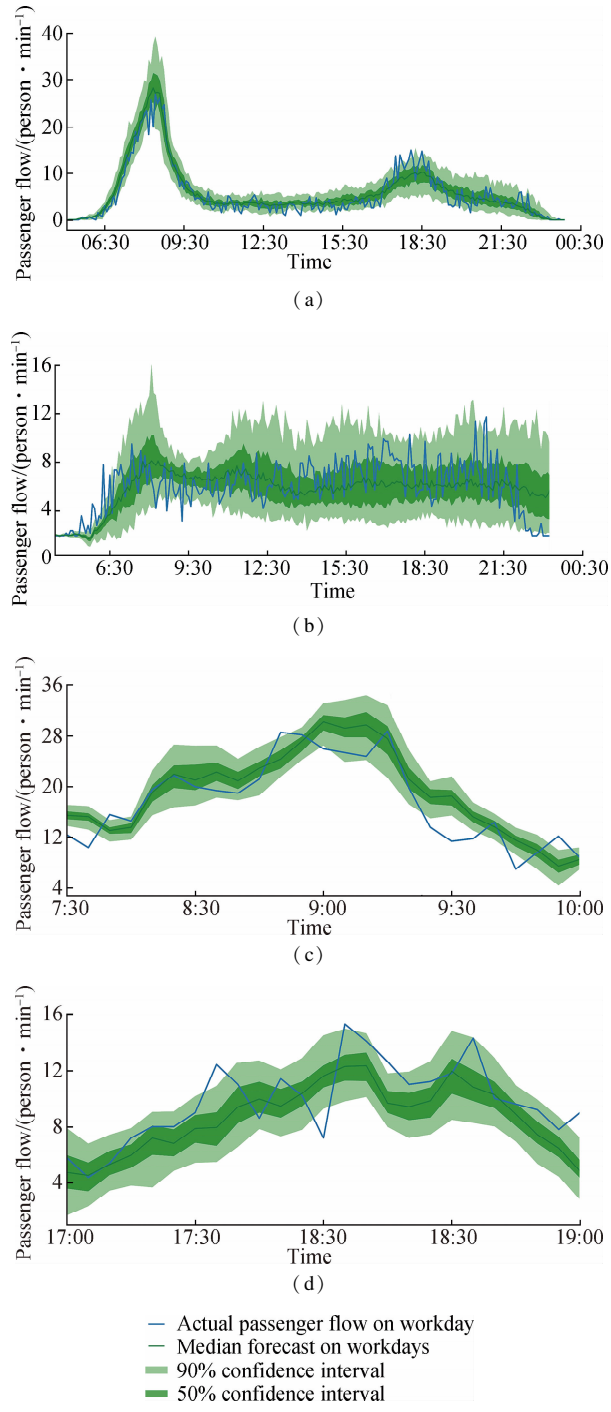

Fig. 5 Workday, non-workday, morning and evening peaks of workday transfer passenger flow forecast. (a) Workday (May 25, 2018); (b) Non-workday (May 26, 2018); (c) Workday (May 24, 2018) morning peak; (d) Workday (May 24, 2018) evening peak

To verify the effectiveness of the PSO-DeepAR model, this study compares six models, namely, Mean, Seasonal Naive, ARIMA, SimpleFeedForward, DeepAR, and PSO-DeepAR. The Mean prediction method is the prediction of all future values equal to the average of historical data. The Seasonal Naive prediction method is based on the characteristics of the data, taking the point of the last cycle of the data as the prediction value, which is a single-point prediction. The SimpleFeedForward neural net-

work is the simplest neural network, where each neuron is arranged in layers, there is no feedback in the whole network, and the signal propagates from the input layer to the output layer in one direction. ARIMA is an autoregressive sliding average method, which transforms the nonstationary time series into a stationary time series and builds a model by regressing the dependent variable on only its lagged values and the present and lagged values of the random error term.

Using six evaluation indicators for assessment, multiple models predicted errors, as shown in Tab. 4.

Tab. 4 Error of multimodel transfer passenger flow prediction

Prediction scenarios	Prediction models	MASE	sMAPE	ND	MSIS	WQL[0.9]	Coverage[0.9]
Weekday	Mean	2.826	0.750	0.723	25.196	0.511	0.872
	Seasonal Naive	0.929	0.353	0.235	36.819	0.254	0.414
	ARIMA	2.805	0.766	0.717	22.300	0.475	0.885
	SimpleFeedForward	0.831	0.394	0.213	6.197	0.092	0.846
	DeepAR	0.840	0.411	0.212	5.108	0.105	0.930
	PSO-DeepAR	0.824	0.425	0.210	4.926	0.214	0.936
Nonworking day	Mean	1.222	0.515	0.422	7.091	0.182	0.837
	Seasonal Naive	0.987	0.397	0.341	39.482	0.400	0.409
	ARIMA	1.068	0.489	0.369	7.030	0.163	0.911
	SimpleFeedForward	0.831	0.450	0.287	5.994	0.160	0.779
	DeepAR	1.063	0.487	0.369	6.447	0.166	0.925
	PSO-DeepAR	1.136	0.476	0.273	6.201	0.165	0.932
Weekday morning peak	Mean	1.631	0.336	0.318	7.392	0.097	0.960
	Seasonal Naive	0.725	0.163	0.141	29.017	0.144	0.520
	ARIMA	1.720	0.361	0.335	7.245	0.085	0.920
	SimpleFeedForward	0.567	0.124	0.110	6.508	0.047	0.880
	DeepAR	0.739	0.157	0.151	3.650	0.082	0.920
	PSO-DeepAR	0.638	0.148	0.139	3.120	0.068	0.934
Weekday evening peak	Mean	1.281	0.276	0.266	5.060	0.080	0.840
	Seasonal Naive	1.000	0.225	0.208	40.001	0.266	0.400
	ARIMA	1.157	0.247	0.241	5.390	0.093	0.720
	SimpleFeedForward	0.834	0.179	0.173	6.197	0.075	0.840
	DeepAR	0.608	0.137	0.135	4.251	0.071	0.870
	PSO-DeepAR	0.574	0.152	0.147	3.972	0.069	0.921

As shown in Tab. 4, the Mean model is the base statistical forecasting model among the five models, and only the Seasonal Naive model is the point forecasting model. The comparison of the point forecasting model and probabilistic interval forecasting model shows that the MASE, sMAPE, and ND errors of the point forecasting model are lower than that of the traditional statistical type probabilistic interval forecasting model and higher than that of the DeepAR model. The MSIS and Coverage[0.9] errors of the point forecasting model are higher than that of the traditional statistical type probabilistic interval prediction model and lower than that of the DeepAR model, i.e., the point prediction model has higher accuracy for MASE but lower accuracy for Coverage[0.9] interval coverage. The DeepAR model is the model with the smallest combined error among several models, and the prediction error of the PSO-DeepAR model is significantly reduced after optimization by the PSO algorithm. Thus, the DeepAR model is the most suitable for predicting transfer passenger flow intervals.

When comparing point forecasts with interval forecasts, the decision-maker should select point forecasts if they need to predict a value or interval forecasts if they need to obtain a range of possible fluctuations in passenger numbers. The value obtained from the point forecast is only a point in the range obtained from the interval forecast. When a metro station needs to cope with a sudden peak in passenger flow, the point forecast does not predict the peak in passenger flow, and the interval forecast is more appropriate.

5 Conclusions

1) Passenger trip chains are identified and extracted according to the transfer time threshold of passenger trips, the specific structure of passenger trip chains is obtained, and the trip chains with a trip phase less than or equal to four at Xizhimen metro station are screened to obtain the passenger flow of the metro-to-bus transfer.

2) An optimized probabilistic interval prediction model is used and compared with the single-point prediction model. The optimized probabilistic interval prediction provides a range of predicted passenger flow fluctuations under a certain confidence level, which improves the accuracy and reliability of the prediction.

3) Four typical scenarios, namely, weekday, non-workday, weekday morning peak, and weekday evening peak, were selected for separate forecasting. The forecasting results of PSO-DeepAR were compared with those of Mean, Seasonal Naive, ARIMA, SimpleFeedForward, and DeepAR. The results show that the PSO-DeepAR model has the highest patronage coverage at a 90% confidence interval and the best interval prediction.

References

[1] Adler T, Ben-Akiva M. A theoretical and empirical model of trip chaining behavior[J]. *Transportation Research Part B: Methodological*, 1979, **13**(3): 243 - 257. DOI: 10.1016/0191-2615(79)90016-X.

[2] Kondo K, Kitamura R. Time-space constraints and the formation of trip chains[J]. *Regional Science and Urban Economics*, 1987, **17**(1): 49 - 65. DOI: 10.1016/0166-0462(87)90068-8.

[3] Tan J M, Xu R H. Analysis of multi-factors influencing trip chain buildup [J]. *Journal of Tongji University (Natural Science)*, 2009, **37**(10): 1340 - 1344. DOI: 10.3969/j. issn. 0253-374x. 2009. 10. 012. (in Chinese)

[4] Qi C, Zhu Z J, Guo X C, et al. Examining interrelationships between tourist travel mode and trip chain choices using the nested logit model[J]. *Sustainability*, 2020, **12**(18): 7535. DOI: 10.3390/su12187535.

[5] Wang C Y, Hu S R, Chu C P. A combined activity nodes choice and trip-chain based user equilibrium traffic assignment model [J]. *Transportation Research Procedia*, 2017, **25**: 2461 - 2472. DOI: 10.1016/j. trpro. 2017. 05. 271.

[6] Wang J C, Chen S K, He Y Q, et al. Simulation of transfer organization of urban public transportation hubs [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2006, **6**(6): 96 - 102. DOI: 10.1016/S1570-6672(07)60004-X.

[7] Wang Z W, Chen T, Song M Q. Coordinated optimization of operation routes and schedules for responsive feeder transit under simultaneous pick-up and delivery mode [J]. *Journal of Traffic and Transportation Engineering*, 2019, **19**(5): 139 - 149. DOI: 10.19818/j. cnki. 1671-1637. 2019. 05. 014. (in Chinese)

[8] Zhang X J. *The study on the transfer between urban rail transportation and conventional public transit* [D]. Chengdu: Southwest Jiaotong University, 2004. (in Chinese)

[9] Xiong J, Guan W, Sun Y X. Metro transfer passenger forecasting based on Kalman filter[J]. *Journal of Beijing Jiaotong University*, 2013, **37**(3): 112 - 116, 121. DOI: 10. 3969/j. issn. 1673-0291. 2013. 03. 021. (in Chinese)

[10] Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model [J]. *Neurocomputing*, 2003, **50**: 159 - 175. DOI: 10.1016/S0925-2312(01)00702-0.

[11] Liu G J, Yin Z Z, Jia Y J, et al. Passenger flow estimation based on convolutional neural network in public transportation system [J]. *Knowledge-Based Systems*, 2017, **123**: 102 - 115. DOI: 10.1016/j. knosys. 2017. 02.016.

[12] Hu Y R, Wu C, Liu H J. Prediction of passenger flow on the highway based on the least square support vector machine[J]. *Transport*, 2011, **26**(2): 197 - 203.

[13] Zhao J D, Gao Y, Bai Z M, et al. Traffic speed prediction under non-recurrent congestion: Based on LSTM method and BeiDou navigation satellite system data[J]. *IEEE Intelligent Transportation Systems Magazine*, 2019, **11**(2): 70 - 81. DOI: 10.1109/MITS. 2019. 2903431.

[14] Zhang X H. *Short-term traffic flow interval forecasting based on grey system theory*[D]. Zhenjiang: Jiangsu University, 2011. (in Chinese)

[15] Zhu S L, Cheng L, Chu Z M. Bayesian network model for traffic flow estimation using prior link flows[J]. *Journal of Southeast University (English Edition)*, 2013, **29**(3): 322 - 327. DOI: 10.3969/j. issn. 1003-7985. 2013. 03.017.

[16] Tong L, Guan Z. Fuzzy granulation prediction of traffic flow based on improved whale optimization support vector machine [J]. *Journal of Computer Applications*, 2021(10): 2919 - 2927. DOI: 10. 11772/j. issn. 1001-9081. 2020122048. (in Chinese)

[17] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: Probabilistic forecasting with autoregressive recurrent networks [J]. *International Journal of Forecasting*, 2020, **36**(3): 1181 - 1191. DOI: 10.1016/j. ijforecast. 2019. 07. 001.

[18] Yan L C, Li Y, Song H, et al. Web traffic prediction based on prophet-DeepAR[J]. *Journal of Guangxi Normal University (Natural Science Edition)*, 2022(3): 172 - 184. DOI: 10. 16088/j. issn. 1001-6600. 2021071505. (in Chinese)

[19] Zhu G, Li W, Du S G, et al. Time series prediction based on deep learning model DeepAR and application examples[J]. *E-Business Journal*, 2020, **39**(7): 83 - 86. DOI: 10. 14011/j. cnki. dzsw. 2020. 07. 039. (in Chinese)

[20] Li H, Wang Z J, Li Z, et al. Prediction of remaining useful life of aero-engine based on stacked Autoencoder and DeepAR [J]. *Journal of Propulsion Technology*, 2022: 1 - 10.

[21] Wang Z J, Zhan Z H, Kwong S, et al. Adaptive granularity learning distributed particle swarm optimization for large-scale optimization [J]. *IEEE Transactions on Cybernetics*, 2021, **51**(3): 1175 - 1188. DOI: 10.1109/TCYB. 2020. 2977956.

[22] Seaborn C W. *Application of Smart Card fare payment data to bus network planning in London, UK* [D]. Massachusetts: Massachusetts Institute of Technology, 2008.

[23] Wang Y Y. *Research on methods of extraction commuting trip characteristic based on public transportation multi-source data* [D]. Beijing: Beijing University of Technology, 2014. (in Chinese)

[24] Özakta H, Kırkavak N, Alpay A N. A paradox of the average waiting time for the case of a single bottleneck on the commuters' route[J]. *Modelling and Simulation in Engineering*, 2021, **2021**: 1 - 9. DOI: 10.1155/2021/2315987.

[25] Yan W P. *Evaluation of public transportation index in*

metropolis [D]. Beijing: Beijing University of Technology, 2012. (in Chinese)

[26] Zhao Y Y, Xia L, Jiang X G. Short-term subway passenger flow prediction model based on empirical modal decomposition and long-term memory neural networks [J]. *Journal of Traffic and Transportation Engineering*, 2020, **20**(4): 194 - 204.

基于出行链特征的地铁换乘公交客流概率区间预测

申 瑾[1] 赵建东[1,2] 高 远[3] 冯迎紫[1] 贾 斌[1]

([1] 北京交通大学交通运输学院, 北京 100044)

([2] 北京交通大学综合交通运输大数据应用技术交通运输行业重点实验室, 北京 100044)

([3] 东北林业大学交通学院, 哈尔滨 150040)

摘要: 为精准分析地铁换乘公交客流的时变差异波动范围,提出应用概率区间预测模型预测换乘客流量. 首先,对公交和地铁数据处理与关联匹配,构建公共交通出行链提取的数据基础. 然后,利用判别换乘关系的合理匹配阈值方法提取公共交通出行链,分析基于出行链的出行基本特征,获取地铁换乘公交客流量. 接着,针对点预测精度不够等问题,提出采用 DeepAR 模型开展区间预测,其中输入为换乘客流量,输出为预测的客流中值和客流区间,预测场景有工作日、非工作日、工作日早高峰以及晚高峰. 其次,为减小预测误差,利用粒子群算法(PSO)优化 DeepAR 模型,构建 PSO-DeepAR 组合模型. 最后,利用北京西直门地铁站数据进行验证. 结果表明,PSO-DeepAR 模型预测准确,90% 置信区间覆盖率最高能达到 93.6%.

关键词: 城市交通;概率区间预测;深度学习;地铁换乘公交客流;出行链

中图分类号: U491.1