Research Article

Fuzzy Temporal Logic Based Railway Passenger Flow Forecast Model

Fei Dou,^{1,2} Limin Jia,³ Li Wang,^{1,3} Jie Xu,³ and Yakun Huang¹

¹ School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

² Subway Operation Technology Centre, Mass Transit Railway Operation Corporation LTD, Beijing 102208, China

³ State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Li Wang; wangli298@sohu.com

Received 10 July 2014; Accepted 9 September 2014; Published 5 November 2014

Academic Editor: Xiaobei Jiang

Copyright © 2014 Fei Dou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Passenger flow forecast is of essential importance to the organization of railway transportation and is one of the most important basics for the decision-making on transportation pattern and train operation planning. Passenger flow of high-speed railway features the quasi-periodic variations in a short time and complex nonlinear fluctuation because of existence of many influencing factors. In this study, a fuzzy temporal logic based passenger flow forecast model (FTLPFFM) is presented based on fuzzy logic relationship recognition techniques that predicts the short-term passenger flow for high-speed railway, and the forecast accuracy is also significantly improved. An applied case that uses the real-world data illustrates the precision and accuracy of FTLPFFM. For this applied case, the proposed model performs better than the k -nearest neighbor (KNN) and autoregressive integrated moving average (ARIMA) models.

1. Introduction

High-speed railway as a kind of large volume passenger transportation mode has been well developed in Europe and Japan and has been developing in China in an even larger scale and has been planned to develop in American continent. In these areas, high-speed railway plays the role of backbone of passenger transportation systems. How to raise operation of the efficiency and how to make the passenger service decision-making more demand-responsive have been the most important focus to the research concerned. As one of the most important basics for the decision-making on high-speed railway transportation pattern and train operation planning, passenger flow forecast is of essential importance, and short-term passenger flow forecast is the key to the success of daily operation management.

Recently, many forecast techniques have been used to solve the prediction problems. Lin and Yang applied the grey forecasting model to forecast the output value of Taiwan's optoelectronics industry accurately from 2000 to 2005 [1]. In [2], four models were developed and tested for the freeway traffic flow forecasting problem. They were

the historical average, time-series, neural network, and non-parametric regression models. The nonparametric regression model significantly outperformed the other models. Du and Ren [3] proposed a prediction model of train passenger flow volume to help the railway administration's analysis of running strategies. The model was analysed based on industrial economic indexes and Cobb-Douglas theory to make the prediction. Particularly, ARIMA model has become one of the most common approaches of parametric forecast since the 1970s. The ARIMA model is a linear combination of time-lagged variables and error terms, which has been widely applied in forecasting short-term traffic data such as traffic flow, travel time, and speed. In [4], time series of traffic flow data are characterized by definite periodic cycles. Seasonal autoregressive integrated moving average (ARIMA) and Winters exponential smoothing models were developed. In [5], it was presented that the theoretical basis for modeling univariate traffic condition data streams as seasonal ARIMA process. In [6], Hamed et al. attempted to develop time-series models for forecasting traffic volume in urban arterials, and the Box-Jenkins ARIMA model turned out to be the most adequate model in reproducing all original time series. As

stated by Brooks, ARIMA performed well and robustly in modeling linear and stationary time series [7]. However, the applications of ARIMA models were limited because they assumed linear relationships among time-lagged variables and they could not capture the structure of nonlinear relationships [8].

The nonparametric regression models have been applied to forecast transportation demand. However, among these nonparametric techniques, KNN method has been rarely adopted in forecast transportation demand. Robinson and Polak proposed the use of the KNN technique to estimate urban link travel time with single loop inductive loop detector data, and the optimized KNN model was found to provide more accurate estimates than other urban link travel time methods [9].

Neural network model has been frequently adopted to predict. In [10], the time-delay recurrent neural network for temporal correlations and prediction and multiple recurrent neural networks were described. And the best performance is attained by the time-delay recurrent neural network. In [11], a hybrid EMD-BPN forecast approach which combined empirical mode decomposition (EMD) and backpropagation neural networks (BPN) was developed to predict the short-term passenger flow in metro systems. In [12], the forecast model of railway short-term passenger flow based on BP neural network was established based on analyzing the principle of BP neural network and time sequence characteristics of railway passenger flow. In [13], a neural network model was introduced that combines the prediction from single neural network predictors according to an adaptive and heuristic credit assignment algorithm based on the theory of conditional probability and Bayes' rule. In [14], Chen and Grant-Muller reported the application and performance of an alternative neural computing algorithm which involves "sequential or dynamic learning" of the traffic flow process. This indicated the potential suitability of dynamic neural networks with traffic flow data. In [15], Li and Chong-Xin employed chaos theory into forecasting. Delay time and embedding dimension are calculated to reconstruct the phase space and determine the structure of artificial neural network, and the load data of Shanxi province power grid of China is used to show that the model is more effective than classical standard BP neural network model.

Support vector machine technique has also been adopted in forecast. In [16], a modified version of a pattern recognition technique known as support vector machine for regression to forecast the annual average daily traffic was presented. Hu et al. utilized the theory and method of support vector machine regression and established the regressive model based on the least square support vector machine. Through predicting passenger flow on Hangzhou highway in 2000–2008, the authors showed that the regressive model of the least square support vector machine had much higher accuracy and reliability of prediction [17].

Since the problem was introduced, high-speed railway passenger flow forecast is vitally important to the organization of high-speed railway. However, several studies have focused on forecasting short-term high-speed railway passenger flow on the basis of the regularity and randomness

FIGURE 1: Daily variation of high-speed railway passenger flow.

of the passenger flow rate. A new method is, therefore, very much needed. Fuzzy temporal logic based passenger flow forecast model (FTLPFFM) is proposed in this paper. Quasi-periodic variation of high-speed railway passenger flow is sufficiently reflected and nonlinear fluctuation of high-speed railway passenger flow is processed using fuzzy logic relationship recognition techniques in the searching process. The proposed model has explicit physical meaning, which reflects variation of high-speed railway passenger flow and has sufficient comprehensibility and interpretability. The characteristics of short-term high-speed railway passenger flow are vitally important to forecast model which is used to improve predictive performance of fuzzy k -nearest neighbor by comparing with other predictive methods in short-term high-speed railway passenger flow forecast.

The remainder of this paper is organized as follows. In Section 2, passenger flow characteristics of the high-speed railway and passenger flow variation in adjacent period are summarized. In Section 3, the change degree of passenger flow is divided into eight grades according to cognitive habit and passenger flow change rate is fuzzified. FTLPFFM is proposed in Section 4. In Section 5, the experiment result for the application of FTLPFFM is compared with ARIMA and KNN models when using three statistics: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). And FTLPFFM appears to be more robust and universally fitting. The last section is the conclusion and future work.

2. Passenger Flow Feature Extraction

In short-term passenger flow forecast, the characteristics of high-speed railway passenger flow are summarized based on time variable because passenger flow has strong correlation to time variable. The data of high-speed railway passenger flow were collected from Beijingnan Railway Station to Jinanxi Railway Station, which is passenger flow in per hour from 26 March to 4 April 2012 (see Figure 1) and daily passenger flow from 14 May to 31 July 2012 (see Figure 2).

Two characteristics of high-speed railway passenger flow are taken into account in FTLPFFM. The first significant characteristic is quasi-periodic which imposes a great impact on passenger flow forecast. The running time of high-speed

FIGURE 2: Weekly variation of high-speed railway passenger flow.

train is between 6:00 and 24:00 and the passenger flow in morning peak and evening peak is more than other periods, which is revealed in Figure 1. Also, the high-speed railway passenger flow is usually stable from Saturday to Wednesday, increases on Thursday, and reaches the peak on Friday, which is revealed in Figure 2. Therefore, the fluctuation cycle of high-speed railway passenger flow is one day and one week. The second one is nonlinear fluctuation which also imposes a great impact on passenger flow forecast. Specifically, the change rate of passenger flow is instable with nonlinear fluctuation for a short time because of many effect factors, such as passengers' income, travel cost, and service quality of transportation, which is revealed in Figures 1 and 2.

3. Regularity of Passenger Flow

Notation:

- $p(t)$: the passenger flow in period t ,
- n : the total number of points of the historical passenger flow series,
- $p(n)$: the current passenger flow state,
- $v(t)$: the passenger flow change rate from $p(t)$ to $p(t+1)$,
- u_i : the interval of passenger flow change rate,
- u'_i : the intermediate value of u_i , $i = 1, 2, \dots, 8$.

The history passenger flow series is denoted by $p(1), p(2), \dots, p(t-1), p(t), p(t+1), \dots, p(n-1), p(n)$. The passenger flow change rates $v(1), v(2), \dots, v(t-1), v(t), v(t+1), \dots, v(n-2), v(n-1)$ between adjacent periods are taken into account, and then the passenger flow change rates are analyzed and variation of passenger flow in adjacent period is summed up.

3.1. Change Rate of Passenger Flow. In order to express passenger flow trend in adjacent period clearly and more accurately, passenger flow change rate is normalized.

Define standardized passenger flow change rate $v(t) = (p(t+1) - p(t))/p_{\max} \in [-1, 1]$, and $p_{\max} = \max(|p(2) - p(1)|, |p(3) - p(2)|, \dots, |p(n) - p(n-1)|)$. For $p(t+1) - p(t) < 0$, the passenger flow descends from period t to $t+1$; for

$p(t+1) - p(t) > 0$, the passenger flow increases from period t to $t+1$; for $p(t+1) - p(t) = 0$, the passenger flow does not change from period t to $t+1$.

In Table 1, the data are collected from Beijingnan Railway Station to Jinanxi Railway Station in Beijing-Shanghai high-speed railway. For example, the maximum value of the passenger flow change in adjacent periods is calculated as $p_{\max} = \max(|p(2) - p(1)|, |p(3) - p(2)|, \dots, |p(n) - p(n-1)|) = 857$; the passenger flow change rate from 8:00–8:30 to 8:30–9:00 on October 10th is calculated as $v(1) = (p(2) - p(1))/p_{\max} = (304 - 70)/857 = 0.273$. Similarly, we can calculate the passenger flow change rates, which are 0.231, 0.5158, -0.8145 , and so forth, as shown in Table 1.

3.2. Variation of Passenger Flow. In order to reveal the regularity of the passenger flow trend clearly and express varying degrees of passenger flow change, respectively, we divide passenger flow change rate into eight intervals applying Zadeh's fuzzy set theory [18].

Define the universe of discourse $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8\}$ and partition it into equal length intervals $u_1 = [-1, -0.75]$, $u_2 = [-0.75, -0.5]$, $u_3 = [-0.5, -0.25]$, $u_4 = [-0.25, 0]$, $u_5 = [0, 0.25]$, $u_6 = [0.25, 0.5]$, $u_7 = [0.5, 0.75]$, and $u_8 = [0.75, 1]$. The midpoints of these intervals are $u'_1 = -0.875$, $u'_2 = -0.625$, $u'_3 = -0.375$, $u'_4 = -0.125$, $u'_5 = 0.125$, $u'_6 = 0.375$, $u'_7 = 0.625$, and $u'_8 = 0.875$. Define fuzzy set A_i based on the redivided intervals; fuzzy set A_i denotes a linguistic value of the passenger flow represented by a fuzzy set, $1 \leq i \leq 8$.

The notations A_1, A_2, A_3 , and A_4 denote that passenger flow decrease is too large, larger, microlarge, and less, respectively. Also, the notations A_5, A_6, A_7 , and A_8 denote that passenger flow increase is less, microlarge, larger, and too large.

Eight membership functions in this paper sufficiently reflect quasi-periodic variation of high-speed railway passenger flow, and the forecast result of FTLPPFM has better accuracy based on eight membership functions. Define the fuzzy membership function of subset A_i , namely,

$$f_{A_1}(x) = \begin{cases} 1, & -1 \leq x \leq -0.75, \\ \frac{-0.5 - x}{0.25}, & -0.75 < x \leq -0.5, \\ 0, & x > -0.5, \end{cases}$$

$$f_{A_2}(x) = \begin{cases} \frac{x - (-1)}{0.25}, & -1 < x \leq -0.75, \\ 1, & -0.75 < x \leq -0.5, \\ \frac{-0.25 - x}{0.25}, & -0.5 < x \leq -0.25, \\ 0, & x > -0.25, \end{cases}$$

$$f_{A_3}(x) = \begin{cases} 0, & x \leq -0.75, \\ \frac{x - (-0.75)}{0.25}, & -0.75 < x \leq -0.5, \\ 1, & -0.5 < x \leq -0.25, \\ \frac{-x}{0.25}, & -0.25 < x \leq 0, \\ 0, & x > 0, \end{cases}$$

TABLE 1: The value of passenger flow, passenger flow change degree, passenger flow change rate, and fuzzy set.

Date	Period	Passenger flow	The value of passenger flow change degree	Passenger flow change rate	Fuzzy set
2011.10.10	7:00–8:00	70	234	0.273	A_6
	8:00–9:00	304	198	0.231	A_5
	9:00–10:00	502	442	0.5158	A_7
	10:00–11:00	944	−698	−0.8145	A_1
	11:00–12:00	246	−222	−0.259	A_3
	12:00–13:00	24	14	0.0163	A_5
	13:00–14:00	38	158	0.1844	A_5
	14:00–15:00	196	−19	−0.0222	A_4
	15:00–16:00	177	21	0.0245	A_5
	16:00–17:00	198	543	0.6336	A_7
	17:00–18:00	741	−563	−0.6569	A_2
	18:00–19:00	178	276	0.3221	A_6
	19:00–20:00	454	467	0.5449	A_7
	20:00–21:00	921	−857	−1	A_1
2011.10.11	7:00–8:00	64	76	0.0887	A_5
	8:00–9:00	140	321	0.3746	A_6
	9:00–10:00	461	559	0.6523	A_7
	10:00–11:00	1020	−839	−0.979	A_1
	11:00–12:00	181	−147	−0.1715	A_4
	12:00–13:00	34	24	0.028	A_5
	13:00–14:00	58	59	0.0688	A_5
	14:00–15:00	117	63	0.0735	A_5
	15:00–16:00	180	−33	−0.0385	A_4
	16:00–17:00	147	508	0.5928	A_7
	17:00–18:00	655	−502	−0.5858	A_2
	18:00–19:00	153	283	0.3302	A_6
	19:00–20:00	436	335	0.3909	A_6
	20:00–21:00	771	−718	−0.8378	A_1
2011.10.12	7:00–8:00	53	97	0.1132	A_5
	8:00–9:00	150	282	0.3291	A_6
	9:00–10:00	432	—	—	—

$$f_{A_4}(x) = \begin{cases} 0, & x \leq -0.5, \\ \frac{x - (-0.5)}{0.25}, & -0.5 < x \leq -0.25, \\ 1, & -0.25 < x \leq 0, \\ \frac{0.25 - x}{0.25}, & 0 < x \leq 0.25, \\ 0, & x > 0.25, \end{cases}$$

$$f_{A_5}(x) = \begin{cases} 0, & x \leq -0.25, \\ \frac{x - (-0.25)}{0.25}, & -0.25 < x \leq 0, \\ 1, & 0 < x \leq 0.25, \\ \frac{0.5 - x}{0.25}, & 0.25 < x \leq 0.5, \\ 0, & x > 0.5, \end{cases}$$

$$f_{A_6}(x) = \begin{cases} 0, & x \leq 0, \\ \frac{x}{0.25}, & 0 < x \leq 0.25, \\ 1, & 0.25 < x \leq 0.5, \\ \frac{0.75 - x}{0.25}, & 0.5 < x \leq 0.75, \\ 0, & x > 0.75, \end{cases}$$

$$f_{A_7}(x) = \begin{cases} 0, & x \leq 0.25, \\ \frac{x - 0.5}{0.25}, & 0.25 < x \leq 0.5, \\ 1, & 0.5 < x \leq 0.75, \\ \frac{1 - x}{0.25}, & 0.75 < x \leq 1, \\ 0, & x > 1, \end{cases}$$

$$f_{A_8}(x) = \begin{cases} 0, & x \leq 0.5, \\ \frac{x - 0.5}{0.25}, & 0.5 < x \leq 0.75, \\ 1, & 0.75 < x < 1. \end{cases}$$

(1)

FIGURE 3: Fuzzified passenger flow change rate.

FIGURE 4: Passenger flow change rate relationships.

Different passenger flow change rates can be fuzzified into corresponding fuzzy sets. For example, as seen in Table 1, the passenger flow change rate from 7:00–8:00 to 8:00–9:00 is 0.273, which is fuzzified to A_6 . The passenger flow change rate from 8:00–9:00 to 9:00–10:00 is 0.231, which is fuzzified to A_5 . The passenger flow change rate from 9:00–10:00 to 10:00–11:00 is 0.5158, which is fuzzified to A_7 . And the passenger flow change rate from 10:00–11:00 to 11:00–12:00 is -0.8145 , which is fuzzified to A_1 . The fuzzification process is depicted in Figure 3. Some fuzzified passenger flow change rates are listed in Table 1.

Fuzzy logic relationships are established by putting two consecutive fuzzy sets, as follows:

$$A_j \rightarrow A_p, A_p \rightarrow A_q, \dots, A_s \rightarrow A_t. \quad (2)$$

“ $A_j \rightarrow A_p$ ” denotes that “the fuzzified passenger flow change rate is A_j from period $t-1$ to t and then the fuzzified passenger flow change rate is A_p from period t to $t+1$ ”.

As seen in Figure 4, the fuzzified passenger flow change rate from 7:00–8:00 to 8:00–9:00 is A_6 and from 8:00–9:00 to 9:00–10:00 is A_5 . Hence, we can establish an fuzzy logic relationship as $A_6 \rightarrow A_5$. Likewise, from Table 1, we can establish the fuzzy logic relationships as $A_6 \rightarrow A_5$, $A_5 \rightarrow A_7$, $A_7 \rightarrow A_1$, $A_1 \rightarrow A_3$, and so forth. Some fuzzy logic relationships are listed in Table 2.

TABLE 2: The fuzzy logic relationship of fuzzified passenger flow change rate.

$A_6 \rightarrow A_5$, $A_5 \rightarrow A_7$, $A_7 \rightarrow A_1$, $A_1 \rightarrow A_3$, $A_3 \rightarrow A_5$,
$A_5 \rightarrow A_5$, $A_5 \rightarrow A_4$, $A_4 \rightarrow A_5$, $A_5 \rightarrow A_7$, $A_7 \rightarrow A_2$,
$A_2 \rightarrow A_6$, $A_6 \rightarrow A_7$, $A_7 \rightarrow A_1$, $A_1 \rightarrow A_5$, $A_5 \rightarrow A_6$,
$A_6 \rightarrow A_7$, $A_7 \rightarrow A_1$, $A_1 \rightarrow A_4$, $A_4 \rightarrow A_5$, $A_5 \rightarrow A_5$
$A_5 \rightarrow A_5$, $A_5 \rightarrow A_4$, $A_4 \rightarrow A_7$, $A_7 \rightarrow A_2$, $A_2 \rightarrow A_6$,
$A_6 \rightarrow A_6$, $A_6 \rightarrow A_1$, $A_1 \rightarrow A_5$, — —

4. Fuzzy Temporal Logic Based Passenger Flow Forecast Model

Notation:

k_i : the number of the passenger flow change rate belonging to A_i ,

k : the size of neighborhood,

d : the dimension of the current passenger flow change rate vector.

4.1. K-Nearest Neighbor Model. The K -nearest neighbor model is one of the most famous pattern recognition statistical models. The KNN model defines neighborhoods as those k cases with the least distance to the input state [19]. The literature indicates that Euclidean distance is usually used to determine the distance between the input state and cases in the database [20]. The predictions can be calculated by averaging the observed output values for cases that fall within the neighborhood when the neighborhood is obtained.

For example, a passenger flow series $p(1), p(2), \dots, p(t-1), p(t), p(t+1), \dots, p(n-1), p(n)$ where n is the total number of points of the series. We search the series to find the nearest neighbors, of the current state $p(n)$. Then, we predict $p(n+1)$ on the basis of those nearest values; for example, if the neighborhood size was $k = 1$ and the nearest passenger flow was $p(t)$, then we would predict $p(n+1)$ on the basis of $p(t+1)$. The value of k in KNN model is more often obtained by empirical analysis. In general, the steps of the KNN model can be listed as follows.

Step 1. Identify the neighborhood size k and the original state of variables.

Step 2. Input all original state of variables into the development database.

Step 3. Calculate Euclidean distance of the current state of variables to each state in development database.

Step 4. Choose output of k -nearest neighborhood on the basis of k shortest Euclidean distance from development database.

Step 5. Calculate the predictive value which is the average of the output of k -nearest neighborhood.

4.2. Fuzzy Temporal Logic Based Passenger Flow Forecast Model. Suppose $P(t) = [p(t), p(t+1), \dots, p(t+d-1), p(t+d)]$ is the t -period historical passenger flow state vector and $V(t) = [v(t), v(t+1), \dots, v(t+d-2), v(t+d-1)]$ is the historical passenger flow change rate vector. For $t = n-d$, $P(n-d)$ and $V(n-d)$ are the current passenger flow state vector and the current passenger flow change rate vector.

4.2.1. Distance Metric. Give the state matrix of passenger flow and the matrix of the passenger flow change rate so as to compare the relationship among the different periods of passenger flow more clearly. The state matrix of passenger flow is given by

$$\begin{bmatrix} P(1) \\ P(2) \\ \vdots \\ P(t-d) \\ \vdots \\ P(n-d) \end{bmatrix} = \begin{bmatrix} p(1) & p(2) & \cdots & p(1+d) \\ p(2) & p(3) & \cdots & p(2+d) \\ \vdots & \vdots & & \vdots \\ p(t-d) & p(t-d+1) & \cdots & p(t) \\ \vdots & \vdots & & \vdots \\ p(n-d) & p(n-d+1) & \cdots & p(n) \end{bmatrix}. \quad (3)$$

The matrix of the passenger flow change rate is given by

$$\begin{bmatrix} V(1) \\ V(2) \\ \vdots \\ V(t-d) \\ \vdots \\ V(n-d) \end{bmatrix} = \begin{bmatrix} v(1) & v(2) & \cdots & v(d) \\ v(2) & v(3) & \cdots & v(1+d) \\ \vdots & \vdots & & \vdots \\ v(t-d) & v(t-d+1) & \cdots & v(t-1) \\ \vdots & \vdots & & \vdots \\ v(n-d) & v(n-d+1) & \cdots & v(n-1) \end{bmatrix}. \quad (4)$$

A common approach to measure the “nearness” in KNN model is to use the Euclidean distance [18]. Therefore, the Euclidean distances of the passenger flow state vectors and the passenger flow change rate vectors are as follows:

$$d_1 = \|P(n-d) - P(t-d)\| = \sqrt{\sum_{j=0}^d [p(n-j) - p(t-j)]^2}, \quad (5)$$

$$d_2 = \|V(n-d) - V(t-d)\| = \sqrt{\sum_{j=1}^d [v(n-j) - v(t-j)]^2}. \quad (6)$$

4.2.2. Forecast Passenger Flow Change Rate. Suppose the neighborhood search procedure identifies k neighbors, the passenger flow state vectors of the k neighbors are $P(t' - d, h) = [p(t' - d, h), p(t' - d + 1, h), \dots, p(t' - 1, h), p(t', h)]$ ($h = 1, 2, \dots, k$) and $p(t' + 1, h)$ is next to $p(t', h)$, and $V(n-d) = [v(n-d), v(n-d+1), \dots, v(n-1)]$ is the current passenger flow state vector. The passenger flow change rates corresponding to $p(t' + 1, h)$ and $p(t', h)$ are $v(t', h) = (p(t' + 1, h) - p(t', h)) / p_{\max}$, $h = 1, 2, \dots, k$. The number of the passenger flow change rate $v(t', h)$ belonging to A_i is k_i , and the value of $v(t', h)$ corresponding to A_i is u'_i .

An approach to forecasting is to compute an average of $v(t', h)$ s of the neighbors that have fallen within the neighborhood:

$$v(n) = \frac{k_1 u'_1 + k_2 u'_2 + k_3 u'_3 + k_4 u'_4 + k_5 u'_5 + k_6 u'_6 + k_7 u'_7 + k_8 u'_8}{\sum_{i=1}^8 k_i}. \quad (7)$$

4.2.3. Steps of FTLPPFM. The establishment of FTLPPFM is based on fuzzy k -nearest neighbor prediction method.

Steps of FTLPPFM are as follows.

Step 1. Start with a minimal neighborhood size, $k = 1$.

Step 2. Start with a minimal dimension of the current passenger flow change rate vector, $d = 1$.

Step 3. Start with period $l = n + 1$ to predict passenger flow.

Step 4 (match to find the elementary neighbors). Find the nearest matches for the current passenger flow state vector $P(l-d-1) = [p(l-d-1), p(l-d), \dots, p(l-2), p(l-1)]$ by searching the passenger flow series $p(1), p(2), \dots, p(n-1)$ using (5), and then sort them in ascending order. Suppose an index $t' - d$, for which the nearest matching passenger flow state vector is $P(t' - d) = [p(t' - d), p(t' - d + 1), \dots, p(t' - 1), p(t')]$ and the historical passenger flow change rate vector associated is $V(t' - d) = [v(t' - d), v(t' - d + 1), \dots, v(t' - 2), v(t' - 1)]$. Here, the current passenger flow change rate vector is $V(l-d-1) = [v(l-d-1), v(l-d), \dots, v(l-3), v(l-2)]$; search the same fuzzy logical relationships $A'_i \rightarrow A'_j \rightarrow \dots \rightarrow A'_p \rightarrow A'_q$ for $V(t' - d)$ and $A_i \rightarrow A_j \rightarrow \dots \rightarrow A_p \rightarrow A_q$ for $V(l-d-1)$, and choose the top $2k$ matches which are the elementary neighbors. The appropriate passenger flow change rate vectors of $2k$ will be discussed below.

Step 5 (match to find the nearest neighbors). Find the nearest matches for $V(l-d-1)$ by searching all the historical passenger flow change rate vectors $V(t' - d)$ using (6), and

then sort them in ascending order and choose the top k matches. They are the nearest neighbor passenger flow state vectors $P(t' - d, h) = [p(t' - d, h), p(t' - d + 1, h), \dots, p(t' - 1, h), p(t', h)]$, and output $p(t', h)$ and $p(t' + 1, h)$, $h = 1, 2, \dots, k$.

Step 6. Estimate the passenger flow change rate $v(l - 1)$ using (7).

Step 7. Calculate predictive value of passenger flow $\bar{p}(l) = p(l - 1) + p_{\max} \cdot v(l - 1)$ and add it to the database; repeat Step 4 to Step 7 with regard to $l = l + 1$ until $l = M$, M is the last period.

Step 8. Calculate RMSE between the actual values and predicted values, which is given by

$$\text{RMSE} = \sqrt{\frac{1}{M - n} \sum_{i=n+1}^M [\bar{p}(i) - p(i)]^2}, \quad (8)$$

where $\bar{p}(i)$ is the predicted value of actual value $p(i)$.

Step 9. Repeat Steps 3 to 8 for vector dimensions of $d + 1, d + 2, \dots, d_{\max}$.

Step 10. Repeat Steps 2 to 9 for neighborhood sizes of $k + 1, k + 2, \dots, k_{\max}$.

Step 11. Choose the optimal predictive values of passenger flow which yields minimal RMSE by optimizing the vector dimensions and the neighborhood.

Choose the maximum dimension of the current passenger flow change rate vector and the maximum neighborhood size according to the characteristics of the passenger flow. Smith and Demetsky (1994) [20] found that the best predictions were generated using $k = 10$, and Karlsson and Yakowitz (1987) [21] proposed that the best forecast values were generated using $k = 3$. Wang et al. (2011) [22] and Oswald et al. (2001) [23] revealed that the best results were obtained when $k \leq 30$. We obtain the best predicted values of passenger flow as nearly all fall within the search space, which is $1 \leq k \leq 30$ and $1 \leq d \leq 20$, by numerous experiments using different dataset.

5. Case Study

The data were obtained from National Key Technology Research and Development Program, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. The database was per hour passenger flow between 7:00 and 21:00 from Beijing to Jinan in Beijing-Shanghai high-speed railway, which was split into two parts separately: an estimation data set and a test data set. The estimation data set was collected from 1 July to 31 December 2011 (2576 observations) and the test data set was collected from 1 to 22 January 2012 (300 observations).

According to the passenger flow characteristics, we can set $d_{\max} = 10$ and $k_{\max} = 20$. The developed model for the

TABLE 3: The comparison between ARIMA, KNN, and FTLPPFM.

Prediction model	MAE	MAPE	RMSE
KNN	5.4623	0.0453	3.5684
ARIMA	2.9642	0.0401	3.1812
FTLPFFM	1.9117	0.0328	2.7046

FIGURE 5: Comparisons of predictive values and real values.

passenger flow of the high-speed railway was implemented using MATLAB version 7.1. The best results were obtained when $k = 10$ and $d = 4$, which can be seen from RMSE performance, and $\text{RMSE} = 2.7046$. The best prediction results and actual values are shown in Figure 5.

ARIMA model is a benchmarking method in forecasting field, but it is a gray box model, which cannot reflect the underlying structural properties. KNN model has dynamic adaptability to the data which is a white box model and has sufficient comprehensibility. And FTLPPFM is presented based on KNN forecasting model and has sufficient comprehensibility and interpretability. Therefore, FTLPPFM is compared with ARIMA and KNN models using three statistics: MAE, MAPE, and RMSE, as is shown in Table 3. And (9) shows how MAE and MAPE are computed, respectively. Consider

$$\begin{aligned} \text{MAE} &= \frac{1}{M - n} \sum_{i=n+1}^M |\bar{p}(i) - p(i)|, \\ \text{MAPE} &= \frac{1}{M - n} \sum_{i=n+1}^M \frac{|\bar{p}(i) - p(i)|}{p(i)}. \end{aligned} \quad (9)$$

The absolute error and the absolute relative deviation of three models are computed as shown in Figures 6 and 7.

The result of the comparison between the prediction results and actual values indicates that the proposed model has been shown to be effective and the error is acceptable.

FIGURE 6: The absolute error of three models.

FIGURE 7: The absolute relative deviation of three models.

6. Conclusion and Future Work

Railway transport is an increasingly popular transportation mode for medium-long distance journey in many countries in recent years. The short-term passenger flow forecast has played a key role in high-speed railway intelligent transportation system. In this paper, a FTLPPFM is developed to measure uncertainty of high-speed railway passenger flow for

railway passenger transport management. In FTLPPFM, the past sequences of passenger flow are considered to predict the future passenger flow using fuzzy logic relationship recognition techniques in the searching process. The results reveal that the forecast accuracy (measured with MAE, MAPE, and RMSE) of the FTLPPFM was significantly better than the accuracy levels of the ARIMA and KNN models. Fuzzy temporal logic based passenger flow forecast model also provides a theoretical foundation in decision-making of resource allocation. In a more general sense of application, the proposed method could be adapted in multimodal transportation systems especially in railway transport and metro transport. For future work, one possible extension of this research is to improve forecast accuracy via properly applying data fusion and pattern recognition techniques.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Project is supported by the National Natural Science Foundation of China (no. 61074151), the National Key Technology Research and Development Program of China (no. 2009BAG12A10), the National High Technology Research and Development Program 863 of China (no. 2012AA112001), and the Research Fund of Beijing Jiaotong University (no. T14JB00380), China.

References

- [1] C.-T. Lin and S.-Y. Yang, "Forecast of the output value of Taiwan's opto-electronics industry using the Grey forecasting model," *Technological Forecasting and Social Change*, vol. 70, no. 2, pp. 177–186, 2003.
- [2] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.
- [3] X. Du and N. Ren, "Railway passenger flow volume prediction model analysis based on Cobb-Douglas function," *Applied Mechanics and Materials*, vol. 178–181, pp. 1961–1964, 2012.
- [4] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transportation Research Record*, no. 1644, pp. 132–141, 1998.
- [5] A. Schadschneider, W. Knospe, L. Santen, and M. Schreckenberg, "Optimization of highway networks and traffic forecasting," *Physica A: Statistical Mechanics and Its Applications*, vol. 346, no. 1–2, pp. 165–173, 2005.
- [6] M. M. Hamed, H. R. Al-Masaeid, and Z. M. Bani Said, "Short-term prediction of traffic volume in urban arterials," *Journal of Transportation Engineering*, vol. 121, no. 3, pp. 249–254, 1995.
- [7] C. Brooks, *Introductory Econometrics for Finance*, Cambridge University Press, Cambridge, UK, 2008.
- [8] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: the state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

- [9] S. Robinson and J. W. Polak, "Modeling urban link travel time with inductive loop detector data by using the k-NN method," *Transportation Research Record*, no. 1935, pp. 47–56, 2005.
- [10] S.-S. Kim, "Time-delay recurrent neural network for temporal correlations and prediction," *Neurocomputing*, vol. 20, no. 1–3, pp. 253–263, 1998.
- [11] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.
- [12] Y. Wang, D. Zheng, S. Luo, D. Zhan, and P. Nie, "The research of railway passenger flow prediction model based on BP neural network," *Advanced Materials Research*, vol. 605–607, pp. 2366–2369, 2012.
- [13] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.
- [14] H. Chen and S. Grant-Muller, "Use of sequential learning for short-term traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 5, pp. 319–336, 2001.
- [15] L. Li and L. Chong-Xin, "Application of chaos and neural network in power load forecasting," *Discrete Dynamics in Nature and Society*, vol. 2011, Article ID 597634, 12 pages, 2011.
- [16] M. Castro-Neto, Y. Jeong, M. K. Jeong, and L. D. Han, "AADT prediction using support vector regression with data-dependent parameters," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2979–2986, 2009.
- [17] Y. Hu, C. Wu, and H. Liu, "Prediction of passenger flow on the highway based on the least square support vector machine," *Transport*, vol. 26, no. 2, pp. 197–203, 2011.
- [18] L. A. Zadeh, "Fuzzy sets and fuzzy systems," *International Journal of General Systems*, vol. 17, pp. 129–138, 1990.
- [19] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.
- [20] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction models—a comparison of neural network and nonparametric regression approaches," *Transportation Research Record*, vol. 1453, pp. 98–104, 1994.
- [21] M. Karlsson and S. Yakowitz, "Rainfall-runoff forecasting methods, old and new," *Stochastic Hydrology and Hydraulics*, vol. 1, no. 4, pp. 303–318, 1987.
- [22] J. Wang, P. Shang, and X. Zhao, "A new traffic speed forecasting method based on bi-pattern recognition," *Fluctuation and Noise Letters*, vol. 10, no. 1, pp. 59–75, 2011.
- [23] R. K. Oswald, W. T. Scherer, and B. L. Smith, "Traffic flow forecasting using approximate nearest neighbor nonparametric regression," Research Report No. UVACTS-15-13-7, 2001.