IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 6, 2019, accepted August 18, 2019, date of publication August 23, 2019, date of current version September 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937114

Short-Term Passenger Flow Prediction in Urban Public Transport: Kalman Filtering Combined K-Nearest Neighbor Approach

SHIDONG LIANG[1], MINGHUI MA[2], SHENGXUE HE[1], AND HU ZHANG[1]
[1]Business School, University of Shanghai for Science and Technology, Shanghai 200093, China
[2]School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Corresponding author: Minghui Ma (maminghui1989@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71801153, in part by the National Natural Science Foundation of China under Grant 71801149, in part by the National Natural Science Foundation of China under Grant 71601118, in part by the Natural Science Foundation of Shanghai under Grant 18ZR1426200, in part by the Shanghai Innovation Training Program under Grant SH2019080, and in part by The Fifth Batch of Applied Undergraduate Programs in Shanghai Municipal Universities "Automobile Service Engineering Major" Construction (Shanghai Education Commission [2017] No. 79).

ABSTRACT Short-term prediction of passengers' flow is one of the essential elements of the operation and real time control for public transit. Although fine prediction methodologies have been reported, they still need improvement in terms of accuracy when the current or future data either exhibit fluctuations or significant change. To address this issue, in this study, a fusion method including Kalman filtering and K-Nearest Neighbor approach is proposed. The core point of this method is to design a framework to dynamically adjust the weight coefficients of the predicted values obtained by Kalman filtering and K-Nearest Neighbor approach. The Kalman filtering and K-Nearest Neighbor approach can handle different variation trend of the data. The dynamic weight coefficient can more accurately predict the final value by giving more weight to the appropriately predicted method. In the case study of real-world data, the predicted values of alighting passengers and boarding passengers are presented by four predicted methods involving Kalman filtering, K-Nearest Neighbor approach, support vector machine, and the proposed method. According to the comparison of the test results, the proposed fusion method performed better in terms of predicting accuracy, even if time-series data abruptly varied or exhibited wide fluctuations. The proposed methodology was found as one of the effective approaches based on the historical data and current data in the area of passengers' flow forecasting for urban public transit.

INDEX TERMS Short-term forecasting, urban public transit, passenger flow, fusion model.

I. INTRODUCTION

Intelligent Transportation Systems (ITS) is a global application deploying significant amount of advanced systems and techniques to solve traffic problems, such as traffic congestion and traffic environment. Obtaining accurate information about the future traffic flow including vehicle intelligent system, vehicle routing, and congestion management has a significant application in ITS. Considering the traffic flow changes with time and the real-time requirement of traffic control, the real time data could not be used as the direct input into the ITS [16]–[20]. Traffic prediction becomes one of the central topics in ITS, fundamentally requiring the advanced and smart transportation technologies. Optimizing traffic prediction performance can provide the prior traffic information for the traffic control and management [30], [32]. Over the past few decades, traffic prediction has attracted significant attention [11], [15], [35], [41], [44], [52]. However, for stochastic characteristics of traffic parameters, accurate traffic prediction is not a straightforward task.

Various traffic techniques have been deployed and widely applied to solve the traffic prediction problem. The existing prediction methods can be generally classified into three groups: the parametric methods, non-parametric methods, and hybrid methods [25]. The parametric methods assume an explicit relationship among the parameters. Typically, the

The associate editor coordinating the review of this article and approving it for publication was Rongni Yang.

parametric method expressions can be descripted by the statistical or physical knowledge of the variable extracted. The parametric methods include the linear autoregressive integrated moving average (ARIMA) method [37], [42], [45], [47], seasonal autoregressive integrated moving average (SARIMA) method [22], and Kalman filtering method [13], [40], [48]. Abadi *et al.* [1] established a short-term traffic flow prediction algorithm using an autoregressive model of the link flows to predict the traffic flow based on the current and historical traffic data in the traffic network. Kumar and Vanajaksh [14] proposed a prediction scheme using the SARIMA model that is particularly relevant to model traffic flow behavior. For discrete-time linear stochastic dynamic systems, Kalman filtering is one of the most widely used recursive algorithms and elegant enough for on-line implementation without complex calculation [51].

Furthermore, the non-parametric method expressions can be directly determined by the available data without any assumptions about data distribution or variable interrelations. The non-parametric methods include the support vector regression (SVR) method [4], [43], [46], Neural network method [22], [28], [34], [38], and K-nearest neighbor (KNN) [3], [8], [12], [23], [31], [49]. Ke *et al.* [10] proposed a two stream multi-channel convolutional neural network (TM-CNN) model for predicting the multi-lane traffic speed. Habtemichael and Cetin [9] proposed a non-parametric and data-driven methodology based on identifying similar traffic patterns using an enhanced K-nearest neighbor approach. Liu *et al.* [24] used the K-nearest neighbor algorithm for short-term traffic flow prediction by improving the distance search method and introducing a multivariate statistical regression model. In contrast with the parametric methods, no explicit choice has to be made about the relationship between the parameters or fitting functions. Although the non-parametric methods have a high accuracy of prediction, these methods always fall easily into slower convergence speed and require a large amount of historical data and time to calibrate the model parameters. Moreover, the hybrid methods integrate the elements of these two approaches by combining the flexibility of the parametric methods and computational efficiency of the non-parametric methods [25], [27], [39]. Although, there are ample reviews on traffic prediction, most of the existing efforts focused on the single historical data or the current data. However, long-range correlations in the traffic data fluctuation persist in the data series. In comparison, traffic flow based on the variety dimension traffic data is hardly predicted [29]. This brings us to the following main objectives to address: To propose a frame of traffic prediction and a novel hybrid prediction method based on the parametric method and non-parametric method considering the current data and the historical data.

Considering time-variety and complexity of passengers' flow data and disadvantages of prediction methods based on the current day data, in this study, a fusion prediction method is proposed based on the current data and the historical database with KNN approach and Kalman filtering method, with the idea of adaptive weights allocation. The proposed prediction method requires both real time data and historical data as the input. KNN and Kalman filtering are used to predict the passengers' flow using the current data and historical data, respectively. The two results are assigned weight coefficients, which can be generated in real time in the process of prediction.

The main contribution and difference with previous works are the proposed framework to dynamically adjust the weight coefficient values according to the real time accuracy of the KNN and Kalman Filter methods, giving more weight to the more accurately predicted value. In addition, in this study, the KNN and Kalman filtering methods were integrated into one framework, assimilating their advantages and eliminating their disadvantages, making their respective advantages complementary to each other. Therefore, the proposed prediction method performs better than only using the KNN or Kalman filtering method.

This paper is organized as follows. Section 2 presents a brief general introduction of the Kalman filtering method and KNN approach used for predicting passengers' flow. Section 3 presents the proposed fusion predicted method based on dynamically adjusting the weight, formulating a fusion model and designing an algorithm framework to obtain the weight coefficient dynamically. A case study based on the real data is described in Section 4 by the KNN method, Kalman filtering method, Support Vector Machine, and the proposed predicting method under different scenarios. Section 5 represents the conclusion and future direction.

II. BASIC MODELS

In this study, traffic patterns are predicted by exploiting similarities. In the proposed fusion method, both the KNN method and the Kalman filtering method are introduced in the framework. The KNN method can predict the passengers' flow using the current day data, and the Kalman filtering method searches the state vector in the historical data. Therefore, in this section, the two traditional prediction methods are introduced, and the related parameters are designed.

A. KALMAN FILTERING METHOD

Kalman filtering, known as a statistically optimal sequential estimation procedure for linear dynamic systems, is an efficient algorithm for needing short series of background information and easily observing any alterations [7]. The Kalman filtering tries to find relationships between some explanatory variables and measures traffic flow data [40]. A short algorithmic procedure description of a classical Kalman filter is provided here.

In Kalman filtering, the state vector $X(t)$ at the secrete time t can be described by the system equation.

$$X(t) = A(t/t-1)X(t-1) + W(t-1). \tag{1}$$

In addition, the observation vector $Z(t)$ connecting to the unknown process is shown by Eq. (2)

$$Z(t) = \mathrm{B}(t)X(t) + \delta(t), \tag{2}$$

where $A(t/t-1)$ and $B(t)$ are the coefficient matrices; $W(t-1)$ is the state noise; and $\delta(t)$ is the observation noise. The state noise $W(t-1)$ and the observation noise $\delta(t)$ are the independent random vectors of the Gaussian.

The Karman filtering method calculates the unknown prediction value by the recursive estimation. Based on the previous time step values, the prediction values are given by

$$\begin{cases} \hat{X}(t/t-1) = A(t/t-1)\hat{X}(t-1) \\ E(t/t-1) = A(t/t-1)P(t-1)A(t/t-1)^T + Q \end{cases} \tag{3}$$

where $\hat{X}(t/t-1)$ is the estimated state of the expected states $X(t/t-1)$ at time t; $E(t/t-1)$ is the filtering error covariance; and Q is the covariance matrix.

Furthermore, the new observation value $Z(t)$ is known. The estimate of $\hat{X}(t)$ at time t is shown by

$$\hat{X}(t) = \hat{X}(t/t-1) + K_g(t)[Z(t) - B(t)\hat{X}(t/t-1)], \tag{4}$$

where $K_g(t)$ is the Kalman gain given by $K_g(t) = A(t/t-1)B(t)^T[B(t)A(t/t-1)B(t)^T + R(t)]^{-1}$; $R(t)$ is the covariance matrix of $\delta(t)$.

The final filtering error covariance $E(t)$ is presented by the following equation.

$$E(t) = [I - K_g(t)B(t)]P(t/t-1)[I - K_g(t)B(t)]^T + K_g(t)R(t)K_g(t)^T \tag{5}$$

Moreover, Eqs. (1)–(5) update the Kalman filtering algorithm from period $t-1$ to t. By the process of prediction and correction, the optimal estimated state is computed based on the minimum covariance estimation.

B. K-NEAREST NEIGHBOR METHOD

Non-parametric algorithms are optimized with a training phase based on the existing data. KNN is a popular non-parametric pattern recognition technique for data mining and statistics owing to its simple implementation and significant classification performance. The KNN can carry out the reliable parametric estimates based on performing discriminant analysis and optimal vector selection, including three components: state vectors, determinant, and predicting function. In this study, KNN is used as the basic algorithm to identify similar traffic profiles.

Suppose there is a database involving traffic flow data. Each time code of the database is one-time series, which also is expressed as a state vector. For example, the database $Y(M, D)$ includes N state vectors consisted by d consecutive time data. In detail, d is the suitable number of lags. The higher the value of d, the more time is required to find the similar vectors. At time t, the vector comprising the traffic flow data in the test database can be written as

$$Y_{h,i}(t) = \left[V_{h,i}(t-1), V_{h,i}(t-2), \cdots, V_{h,i}(t-d)\right], \tag{6}$$

where $Y_{h,i}(t)$ is the vector from the database $Y(M, D)$ and $V_{h,i}(t)$ is the traffic flow data during the sample interval. According to the basic principle Eq. (6), the traffic data in the database $Y(M, D)$ is denoted as

$$Y(M, D) = \begin{bmatrix} Y_{h,1}(d+1), Y_{h,1}(d+2), \cdots, Y_{h,1}(D-1), Y_{h,1}(D) \\ Y_{h,2}(d+1), Y_{h,2}(d+2), \cdots, Y_{h,2}(D-1), Y_{h,2}(D) \\ \vdots \quad \vdots \quad \cdots \quad \vdots \quad \vdots \\ Y_{h,M}(d+1), Y_{h,M}(d+2), \cdots, Y_{h,M}(D-1), Y_{h,M}(D) \end{bmatrix} \tag{7}$$

where M is the horizontal time dimension; D is the end time of the horizontal time dimension.

In order to accurately predict the traffic flow data, the real-time traffic state vector $P(t)$ can be written as

$$P(t) = \left[\hat{V}(t), V(t-1), \cdots, V(t-d-1), V(t-d)\right], \tag{8}$$

where $\hat{V}(t)$ is the prediction traffic flow. The traffic flow data of the real-time traffic state vector $P(t)$ can be presented in Fig. (1) that depicts the process with time.

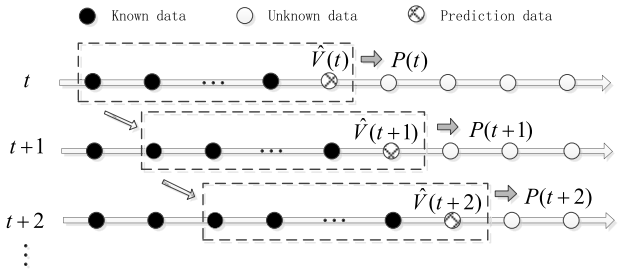


FIGURE 1. Dynamic real-time traffic state vector.

In order to estimate the future traffic flow data, the correlation between the selection vector and the prediction vector is measured by Euclidean distance. Based on the calculated distance, the KNN vectors, collected as the KNN vector group, are achieved.

$$L_{(y_i,p)} = \sqrt{\begin{matrix} \left(V_{h,i}(t-1) - V(t-1)\right)^2 \\ + \left(V_{h,i}(t-2) - V(t-2)\right)^2 \\ + \cdots + \left(V_{h,i}(t-d) - V(t-d)\right)^2 \end{matrix}} \tag{9}$$

Then, the matched data in the KNN vector group is selected to further calculate the prediction data. Furthermore, the prediction result of the KNN method can be written by weighting according to

$$\hat{V}(t) = \sum_j w_j V_{h,j}(t) \tag{10}$$

where $V_{h,j}(t)$ $(j = 1, 2, \cdots, N_k)$ belongs to the KNN vector group; N_k is the number of the KNN vectors; and w_j is the weight coefficient obtained by Eq. (11).

$$w_j = \frac{\sum_{a=1}^{N_k} l_{(y_a,p)} - l_{(y_j,p)}}{(N_k - 1)\sum_{a=1}^{N_k} l_{(y_a,p)}} \tag{11}$$

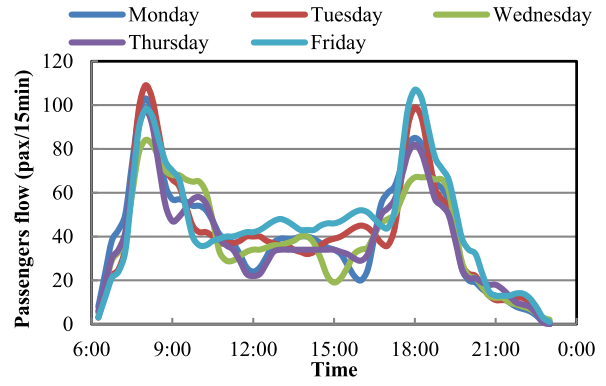
where $l_{(y_j,p)}$ is a vector of the KNN vector group.

III. THE FUSION PREDICTED METHOD BASED ON ADJUSTING WEIGHT DYNAMICALLY

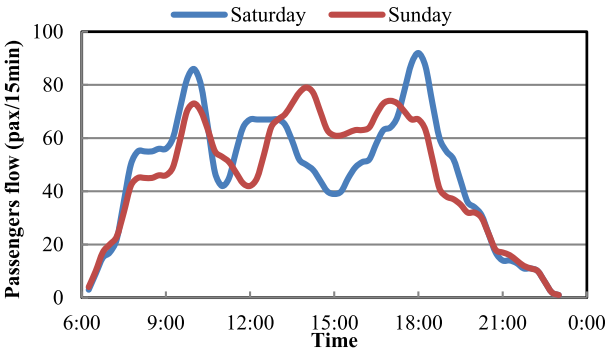
In this section, the characteristic of passengers' flow data is analyzed based on a typical week, followed by formulating the fusion model. The main contribution of this study is designing an algorithm framework to dynamically adjust the weight coefficient values, as presented in Subsection 3.3.

A. DATA FOUNDATION USED IN THE PROPOSED METHOD

Generally, the characteristics of passengers' flow in the urban public transit in weekday and weekend are different. In weekday, the public transit mainly provides service for the commuting passengers, and the trips in weekday are mainly to reach the workplace or for business. In weekend, more passengers would like to go shopping or for relaxing by public transit, and the trips in weekend are mainly for entertainment or relaxation. Therefore, the distribution and total passengers' flow are different between weekday and weekend, as graphically shown in Fig. 2.



(a) Weekday



(b) Weekend

FIGURE 2. Passengers' flow data in one week.

Fig. 2 shows the passengers' flow data in one week (from Monday to Sunday), basically indicating that there are two peak periods in one day including morning peak period and evening peak period. Between these two peak periods, the passengers' flow values are relatively small, as shown in Fig. 2(a). However, the peak hours are not quite obvious in weekend, and the time of peak flow delays in weekend (at 10:00 am) as shown in Fig. 2(b). Therefore, the distribution of passengers' flow in weekday is similar to each other and in weekend is similar with each other. Based on the analysis above, one week can be treated as a cycle. Although the flow data are similar in weekday or in weekend, the predicted values using the historical data cannot be obtained on the same day in the last several weeks, because there may be emergency or activity happened causing exceptional data. Therefore, in the proposed prediction method, the KNN method is introduced into the framework to search for the similar data in the historic database and thus can improve the accuracy of the predicted method. When the passengers' flow is predicted, the KNN method can search in the database avoiding selecting the appropriate data in certain days manually. The Kalman filtering method can predict the values using the current day data. Therefore, these two predicted methods can use the whole database including historical and current day data. The two predicted values were further integrated to obtain the final predicted value using an appropriate method, and this is the core idea of this study, as discussed in the next two subsections.

B. FUSION MODEL FOR THE PROPOSED METHOD

The two predicted values of passengers' flow $\hat{q}_{kf}(t)$ and $\hat{q}_{knn}(t)$ are obtained by the Kalman filtering method and KNN method, respectively. A fusion model was formulated to conveniently calculate the fusion result finally. There are two requirements for the formulation of fusion model. First, the fusion function should be elegant enough to be obtained in real time. Because the Kalman filtering method does not work very well when the data fluctuate greatly, the second requirement is that the weight coefficient parameter $w(k)$ in the fusion model should be predicted using the Kalman filtering with satisfactory accuracy. Based on these two requirements, the proposed fusion function is as follows.

As mentioned above, the fusion model has two predicted values including $\hat{q}_{kf}(t)$ and $\hat{q}_{knn}(t)$. The common concept is to give two weight coefficient values to the two predicted values, which can be written as Eq. (12).

$$\hat{q}(t) = \hat{w}_{kf}(t)\hat{q}_{kf}(t) + \hat{w}_{knn}\hat{q}_{knn}(t) \tag{12}$$

In this fusion model, the two weight coefficients should be calculated. However, the value of passenger flow in the next time $\hat{q}(t+1)$ should be predicted and the real value $q(t)$ in current time has been known. The values of weight coefficient $w_{kf}(t)$ and $w_{knn}(t)$ cannot be calculated by one formula. Therefore, a single weight coefficient fusion model can be formulated, as presented by Eq. (13).

$$\hat{q}(t) = \hat{w}(t)\hat{q}_{kf}(t) + \hat{q}_{knn}(t). \tag{13}$$

If the real passengers' flow value at time t is known, and the real value of the weight coefficient can be calculated by

the inverse function of Eq. (13), written as Eq. (14).

$$w(t) = \frac{q(t) - \hat{q}_{knn}(t)}{\hat{q}_{kf}(t)}. \tag{14}$$

In addition, the weight coefficient value for $\hat{q}_{kf}(k)$ can be set as 1. Therefore, the relationship of Eq. (12) can be expressed by

$$\hat{q}(t) = \hat{q}_{kf}(t) + \hat{w}(t)\hat{q}_{knn}(t). \tag{15}$$

The weight coefficient of $w(t)$ can be calculated by Eq. (16)

$$w(t) = \frac{q(t) - \hat{q}_{kf}(t)}{\hat{q}_{knn}(t)}. \tag{16}$$

Because the time interval of passengers' flow collection on a public transit is generally set as 5 min, 10 min or 15 min, and the passengers' flow is not quite small even during the flat peak periods, zero flow value unlikely occurs. Using the fusion model proposed in this section, no values or pretreatment should be set in advance, indicating convenience in applications.

C. ALGORITHM FRAMEWORK TO ADJUST WEIGHT DYNAMICALLY

As mentioned in Subsection 2.2, the KNN approach can search for the similar values in the database. The Kalman filtering method can predict the value of passenger flow in the next time interval based on the current day data. Using the KNN approach and Kalman filtering method, two predicted values can be obtained. Therefore, in Section 3.2, the fusion model is proposed to integrate the two values obtained by these two methods into a more accurate predicted value. However, as shown in Eq. (12), The fusion model has a weight coefficient. In general, the weight coefficient can be calibrated with the historical data as the mean value. However, this type of calibration method is not quite sensitive to the changing of real time data.

For example, if the data in current day are not impacted by a great disturbance, passengers flow fluctuation is relatively smooth, and both the KNN and Kalman Filter methods can obtain the predicted values with similar accuracy. However, if a non-recurrent social activity is held in the city, historic passengers flow data have less reference value than the current day data. Therefore, the weight coefficient should be calibrated or obtained dynamically, instead of a constant value.

According to the analysis mentioned above, in this study, we propose a dynamic calibration algorithm framework, which can dynamically adjust the weight coefficients to predict the fusion model more accurately. The core concept of the framework is to predict the weight coefficient using the Kalman filtering method again.

As shown in Fig. 3, using the Kalman filtering method based on the passengers' flow data in the current day, the flow data $\hat{q}_{kf}(t)$ was predicted. Meanwhile, the other predicted flow value $\hat{q}_{knn}(t)$ was obtained by the KNN method based on the historical database. The final predicted flow data $\hat{q}(t)$

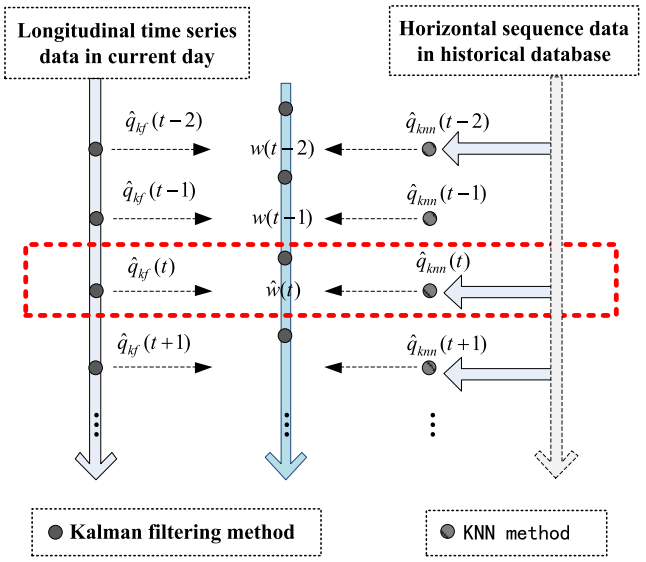


FIGURE 3. The mechanism to dynamically obtain weight coefficient.

can be calculated based on Eq. (12). However, the weight coefficient is unknown. Here, the Kalman filtering method is used again to obtain the $\hat{w}(t)$ based on the historical data of the weight coefficient $w(t-1)$, $w(t-2)$, $w(t-3)$. According to this type of recurrence formula, the predicted weight coefficient value $\hat{w}(t)$ can be obtained continuously.

The more detailed procedure of the proposed predicting method is shown as follows.

Step 1: Using the Kalman filtering method to predict the passenger flow $\hat{q}_{kf}(t)$ at time t based on the passengers flow data in current day.

Step 2: Using the KNN method to predict the value of passenger flow $\hat{q}_{kf}(t)$ at time t based on the historical database of passengers' flow.

Step 3: Calculate the historical values of the weight coefficient in the current day by the inverse function of the fusion model Eq. (14), based on the predicted values.

$\hat{q}_{kf}(t-1)$, $\hat{q}_{kf}(t-2)$, $\hat{q}_{kf}(t-3)$, $\cdots$ and
$\hat{q}_{knn}(t-1)$, $\hat{q}_{knn}(t-2)$, $\hat{q}_{knn}(t-3)$, $\cdots$.

Step 4: Predict the weight coefficient $\hat{w}(t)$ in the fusion model of Eq. (14) by the Kalman filtering method based on the historical values of the weight coefficient in the current day $w(t-1)$, $w(t-2)$, $w(t-3)$, $\cdots$ calculated in Step 3.

Step 5: Calculate the final predicted value of passenger flow $\hat{q}(t)$ at time t using Eq. (12), based on the three predicted values including $\hat{q}_{kf}(t)$, $\hat{q}_{knn}(t)$, and $\hat{w}(t)$ obtained in Steps 1, 2, and 4, respectively.

The more indicative flow chart of the proposed predicting method is shown in Fig. 4. According to Fig. 4, the predicted value of passenger flow value can be obtained at time t. Using this method framework continuously, the real time passenger flow values can be predicted in current day.

IV. CASE STUDY

There are four subsections in the case study including the performance indexes selection, experimental design for the

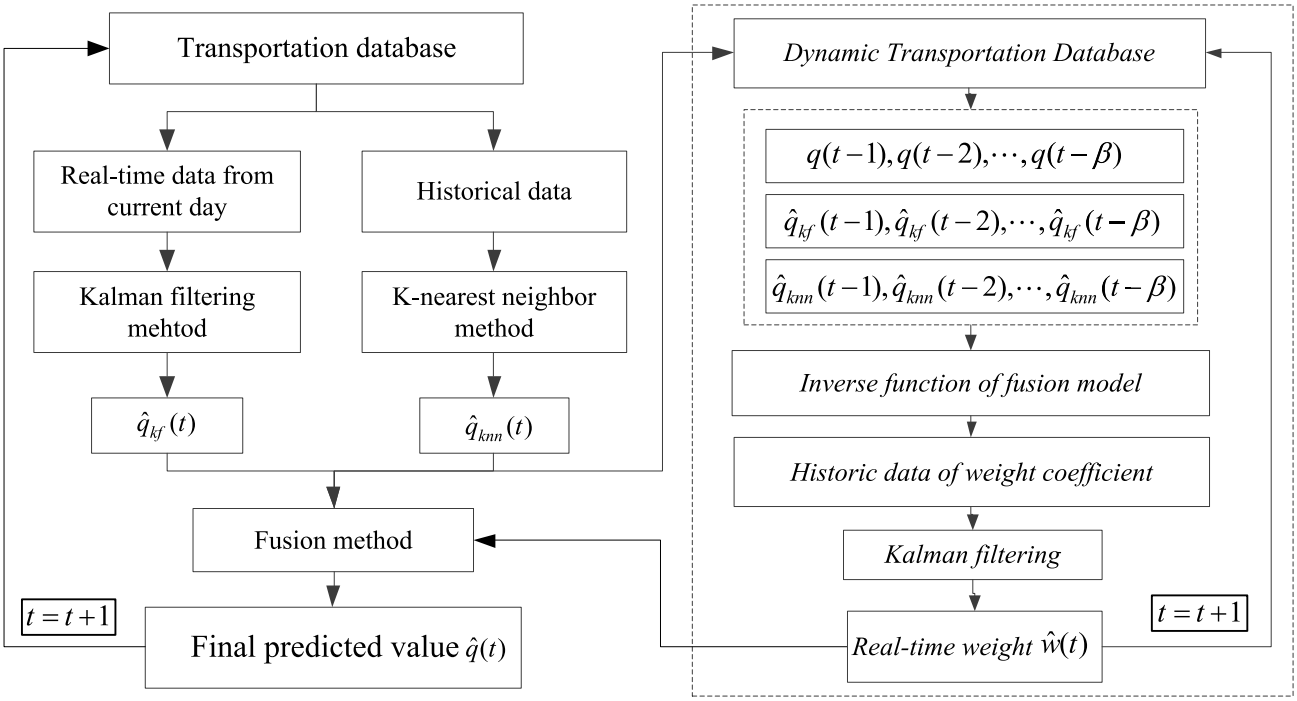


FIGURE 4. The flow chart to predict the passenger flow value at time *t*.

case study, comparison of the results of different predicting methods, and the analysis of the test results.

A. PERFORMANCE EVALUATION

The efficiency of the fusion prediction method for passengers' flow of public transit is evaluated by the mean relative error (MRE) and root mean square error (RMSE). MRE expressed by Eq. (17) indicates the expected error as a fraction of the measurement, providing the error in terms of the percentage of the difference between the real and predicted data values.

$$MRE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{q(t)-\hat{q}(t)}{q(t)}\right| \times 100\%, \tag{17}$$

where $q(t)$ is the real value at time t; $\hat{q}(t)$ is the predicted value at time t; and n is the number of predicted values.

RMSE, expressed by Eq. (18), is the arithmetic mean of the squares of a set of difference between the real values and the predicted values, penalizing large prediction errors.

$$RMSE = \frac{1}{n}\sqrt{n\sum_{t=1}^{n}(q(t)-\hat{q}(t))^2}, \tag{18}$$

B. EXPERIMENTAL DESIGN

Field data are required for the parametric and non-parametric methods to be trained and calibrated, resulting in a fusion predicted method. This study selects Line 3 of the Light Rail Transit in the city of Changchun, China, as an example

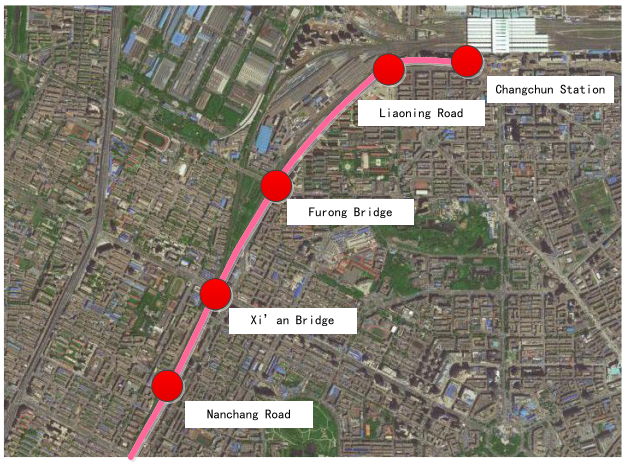


FIGURE 5. The selected stations on Line 3 of light rail transit in city of Changchun.

to test the performance of the proposed predictive method for the passengers' flow at the stations in public transit. The passengers' flow at the five stations, shown in Fig. 5, is collected based on the IC (Integrated Circuit Card) for both getting on and getting off passengers. These five stations are different in terms of average passenger volume and temporal distribution. Changchun Station is located at the Changchun railway station, and the other four stations are mainly for commuting of the dwells.

The passengers' flow data of both getting on and getting off passengers at the five selected stations in March 2017

were collected. In the test, the passengers' flow data on March 31, 2017 are assumed unknown. Therefore, the predicted values of passengers' flow can be compared to the real data values. The time interval of the test is set as 15 min. The operating time of Line 3 is from to 6:00 to 23:00, 17 h in one day. Therefore, there are 68 data values in the test. There are four predictive methods implemented in the test involving Kalman filtering method used only (KF), K-Nearest Neighbor Approach used only (KNN), Support Vector Machine (SVM), and the proposed integrated predicting method (KK) with KF and KNN.

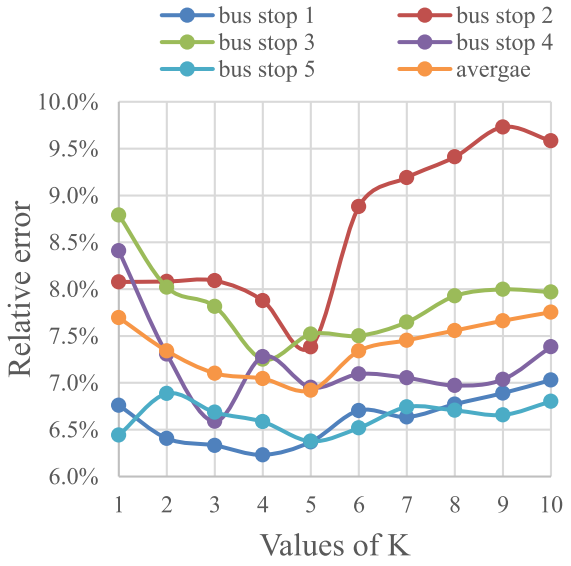


FIGURE 6. Effects of the k-values on the prediction accuracy.

Fig. 6 shows the effects of the k-values on the forecasting accuracy according to the different k-values (from 1 to 10) at the five bus stops using the data in current month. The MAPE decreases to the minimum, and then gradually increases with slight variation between k-values 3 and 6. According to the orange line in Fig. 6, representing the average relative error, the optimal k-value was identified as 5 for this case study.

C. COMPARISON OF RESULTS

The predicted results for boarding and alighting passengers in the test are presented separately. As two typical stations representing the station with high passenger volume at traffic hinge and providing services for the commuting passengers, the predicted values based on the KF, KNN, SVM, and KK methods at Changchun Station and Furong Bridge Station are presented in Figs. 7 and 9. In order to better illustrate the performance of these four predicting methods, the absolute error for these two bus stops is presented in Figs. 8 and 10.

In Fig. 7, the x-axis means the time where one point refers to 15 min. The y-axis represents the number of arrived passengers or left passengers in 15 min. As shown in Fig. 7, the arrived passengers' flow and the left passengers' flow

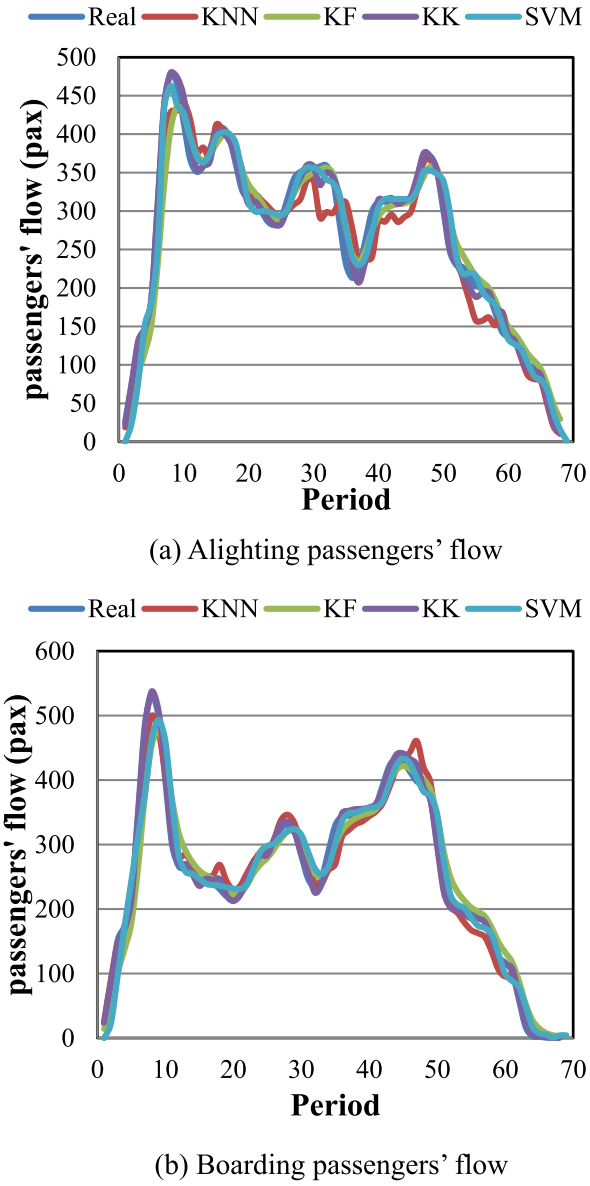
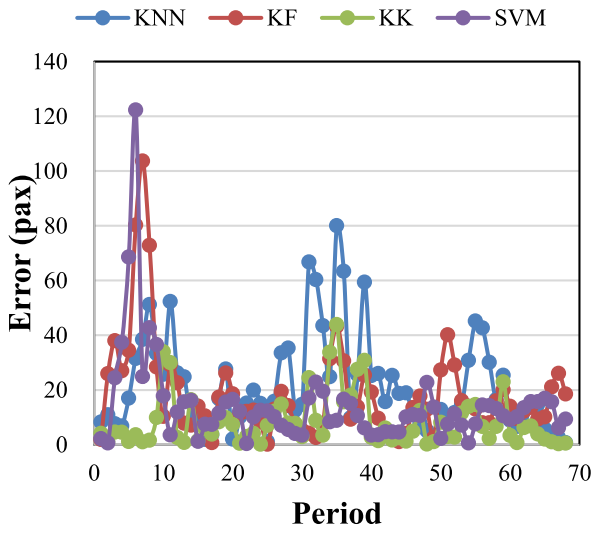


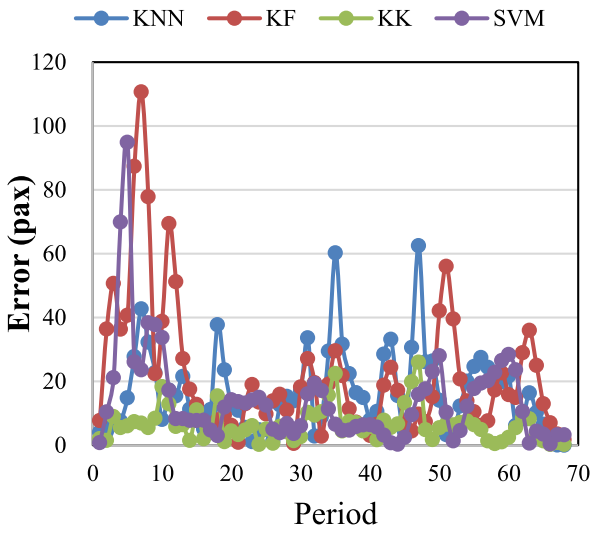
FIGURE 7. Comparison of predicted passengers' flow at Changchun station.

fluctuate at a high level. The phenomenon of peak hours shown in Fig. 7(a) is not very obvious. In addition, as shown in Fig. 7(b), the boarding passengers' flow during evening peak hours is higher than alighting passengers' flow shown in Fig. 7(a), because the passengers tend to leave the city by train in the morning instead of the afternoon.

The blue line in Fig. 7 indicates the real values of passengers' flow data; the red line refers to the predicted values of passengers' flow data by the KNN method; the green line means values of passengers' flow data by the KF method, the purple line represents the predicted values of passengers' flow by the proposed fusion method combined with the KNN with KF and the other line represents the predicted passengers' flow by the SVM method.

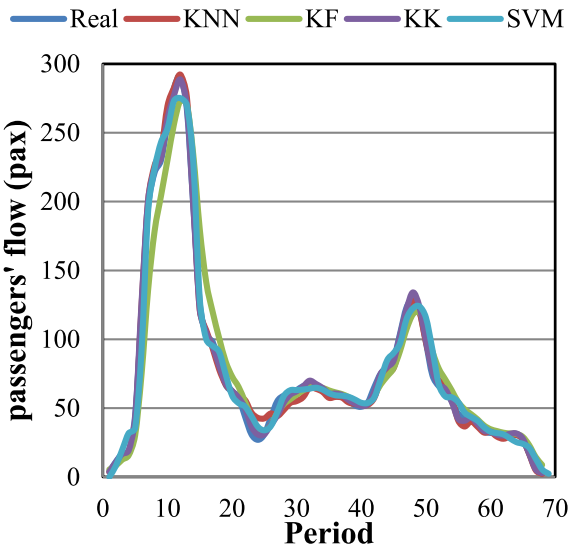


(a) Alighting passengers' flow

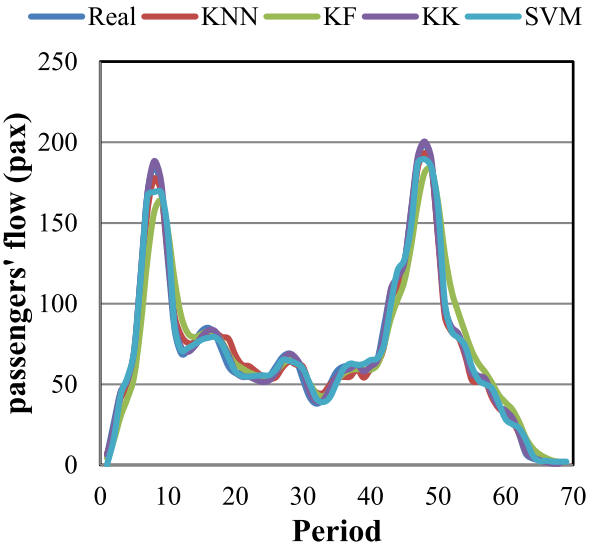


(b) Boarding passengers' flow

FIGURE 8. Performance comparison of absolute errors by the predicted method at Changchun station.



(a) Alighting passengers' flow



(b) Boarding passengers' flow

FIGURE 9. Comparison of the predicted passengers' flow at the Furong Bridge station.

Fig. 7 shows that the variation tendency of purple line is more similar with the blue line, indicating that the KK method has more accuracy than the other three methods, especially during the period flow data fluctuating significantly, as shown in Fig. 7(a) from period 25 to 35.

Fig. 8 provides more details about the performance of the four predictive methods in terms of absolute errors. According to Fig. 8, the KF and SVM methods perform worst during the peak hours, while the KNN method cannot deal with the off-peak hours very well; however, the proposed KK method maintains a low error level, less than 20 pax. Therefore, according to the preliminary impression, the KK method performs better than the other three methods.

Fig. 9 shows the passengers' flow at the Furong Bridge Station. During the morning peak hours, the flow of alighting passengers is relatively high with about 300 passengers for per 15 min, while the number of boarding passengers during the morning peak hours is less than 200. In contrast, the flow of boarding passengers is larger than the number of alighting passengers, because the dwellers should return to their home after getting off the work. During the off-peak period, the passengers' flow data fluctuate at a relatively low level.

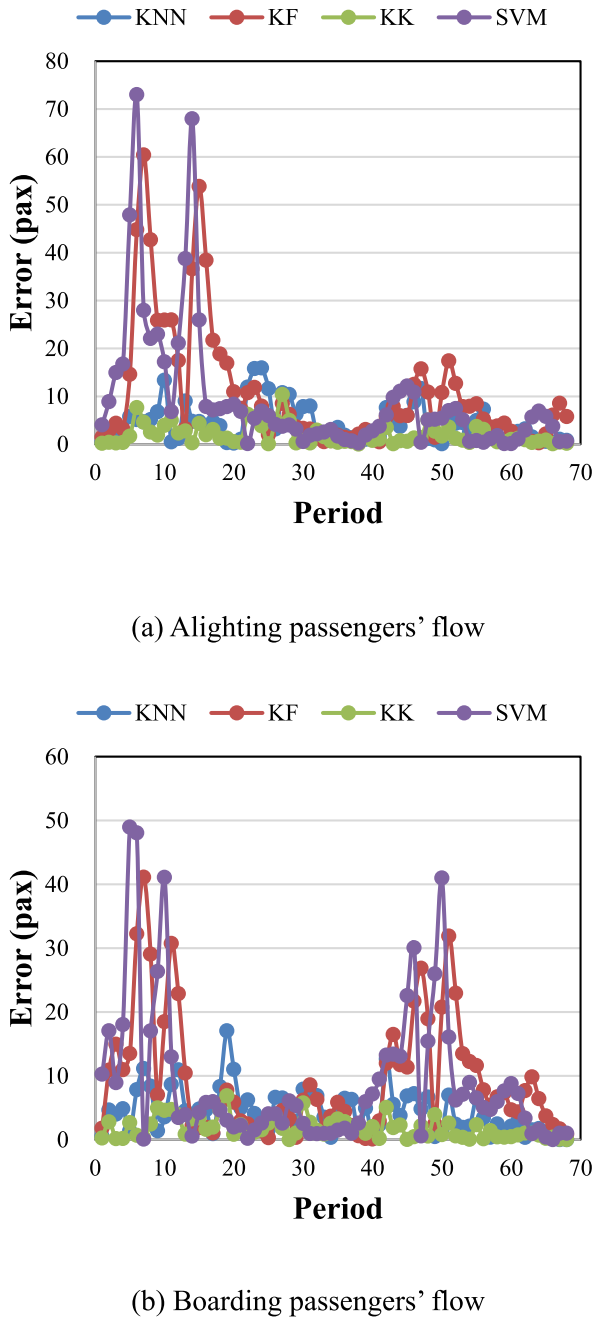


FIGURE 10. Performance comparison of absolute errors for predicted method at Changchun station.

The five colored lines in Fig. 9 represent the same as those in Fig. 7. As shown in Fig. 9(a), during the morning peak hours, the flow data values increased greatly in a short period. The green line does not match the blue line very well, indicating that the KF method performs worse than the other predicted methods when the data changes greatly, and the SVM method has the same inadequacy during the peak hours. In addition, during the off-peak hours, the red line cannot match the blue line very well, indicating that the KNN method performs worse than the other three predicted

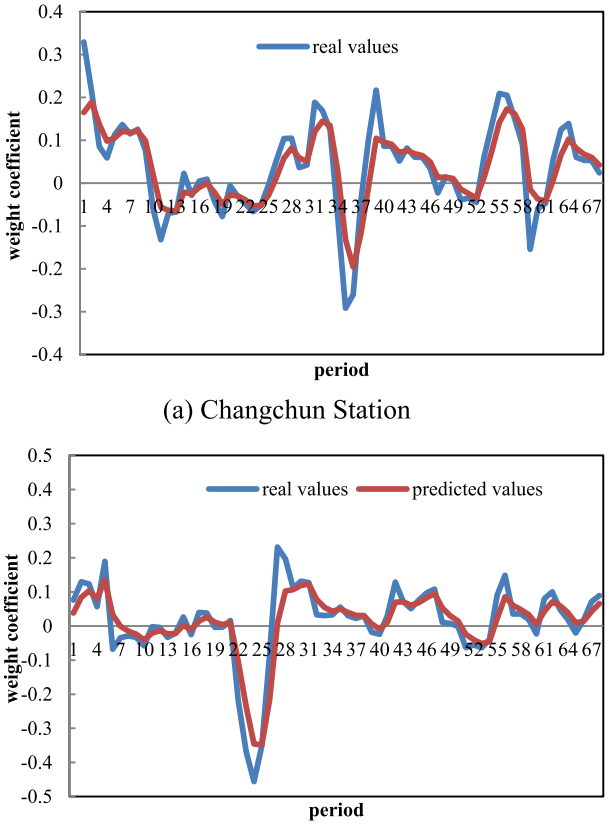


FIGURE 11. Weight coefficient for alighting passengers' flow.

methods when the data fluctuate frequently. The proposed KK method performs better than the other three methods.

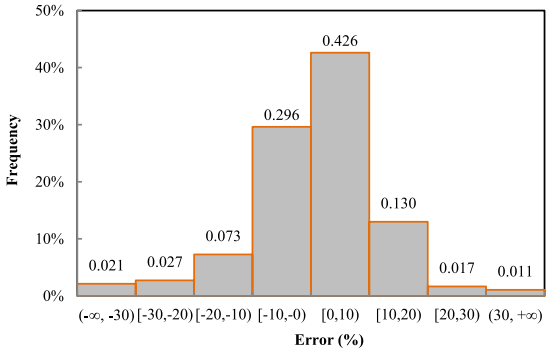
Fig. 10 further shows the characteristics of the four prediction methods when the data have different features. The KF and SVM methods have high absolute errors when the passengers' flow values are large and change greatly, while the other two methods perform better, and this conclusion is similar to that obtained by Fig. 8.

In order to further reflect the detailed performance of the four predicting methods, the index values are calculated and presented in the following subsection.

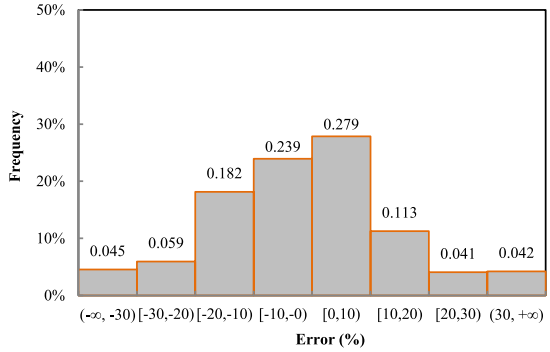
D. ANALYSIS OF THE TEST RESULTS

The core of the proposed predicting method is to obtain the weight coefficient values dynamically. Fig. 11 shows the weight coefficients for alighting passengers' flows at the Changchun Station and Furong Bridge Station.

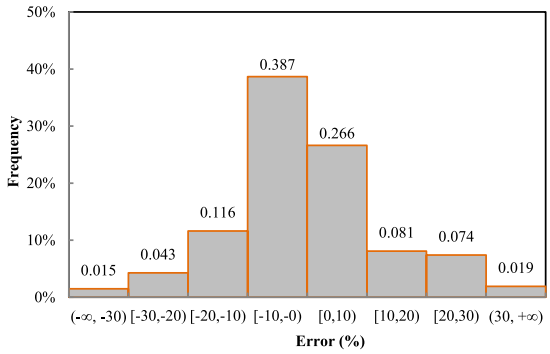
As shown in Fig. 11, the weight coefficient can change dynamically according to the real time accuracy of the KNN and KF methods. The real values of weight coefficient fluctuate slightly from −0.1 to 0.1, except for several extreme points. The predicted weight coefficient values using the KF method can match the real values very well, indicating that the fusion model is formulated appropriately avoiding significant change in the weight coefficient values, and the KF method



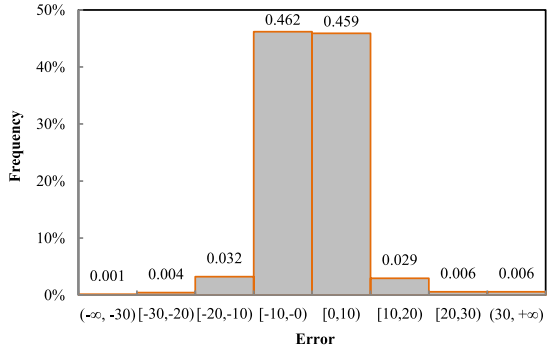
(a) KNN method



(b) KF method



(c) SVM method



(d) KK method

FIGURE 12. Distributions of MRE by the three predicted methods.

TABLE 1. MRE index values by the four prediction methods.

MRE		bus stop 1	bus stop 2	bus stop 3	bus stop 4	bus stop 5
Alighting	KNN	9.0%	7.9%	7.7%	8.3%	7.3%
	KF	7.7%	11.8%	10.9%	13.5%	11.8%
	SVM	6.5%	8.8%	8.8%	12.3%	7.8%
	KK	3.6%	4.0%	4.2%	3.3%	3.4%
Boarding	KNN	6.7%	8.2%	9.9%	8.1%	6.6%
	KF	9.9%	13.9%	14.7%	13.2%	15.5%
	SVM	12.0%	9.9%	9.8%	11.8%	12.5%
	KK	7.6%	4.6%	4.7%	3.7%	2.9%

is competent to predict the weight coefficient in the fusion model. To further reflect the performance of the four prediction methods, the MRE index is presented in Table 1.

In Table 1, the bus stops 1, 2, 3, 4, and 5 refer to Changchun Station, Liaoning Station, Furong Bridge Station, Xi'an Station and Nanchang Station, respectively. The values in Table 1 indicate the MRE of the predicted passengers' flow at the bus stops by different predicting methods. The first, second, third, and fourth row represent the errors for predicting alighting passengers' flow by the KNN, KF, SVM, and KK methods, respectively, while the last four rows correspond to the boarding passengers' flow.

In Table 1, the values in fourth and eighth rows are smaller than the error values in other rows, indicating that the proposed KK predicted method performs better than the other three methods. Most MRE values obtained by the KK method fluctuate around less than 5%, whereas the values by the KNN, KF, and SVM methods fluctuate around 8%, 12%, and 11%, respectively. Furthermore, the distributions of MRE by the KNN, KF, SVM, and KK predicted methods are presented in Fig. 12.

In Fig. 12, the x-and y-axes represent the range of error and the frequency located in the error range, respectively. Note that the relative error calculated here is not an absolute value. Comparing Figs. 12(a), (b), (c), and (d), the hit rates, in terms of MRE, are 72.2, 51.8, 65.3, and 92.1 within 10% and 92.4, 81.2, 85.0, and 98.2 within 20%, respectively. Thus, the proposed KK method more accurately predicted the values of passengers' flow, with 10% errors, which is quite better than the other three methods.

The bus stop numbers in Table 2 are the same as listed in Table 1. The values in Table 2 refer to the RMSE of the predicted values by the four predicting methods. According to Table 2, the RMSE values at bus stop 1 are larger than those at the other four bus stops, because the passengers' flow at bus stop 1 is the largest. In addition, the obtained RMSE values by the proposed KK method are the smallest among the four prediction methods. The KNN method performs

TABLE 2. RMSE index for passengers' flow prediction by different methods.

RMSE (pax)		bus stop 1	bus stop 2	bus stop 3	bus stop 4	bus stop 5
Alighting	KNN	27.31	3.39	2.52	5.81	2.22
	KF	25.84	6.49	5.32	16.97	7.13
	SVM	22.39	4.97	3.92	16.84	4.65
	KK	12.79	1.54	1.19	2.83	1.10
Boarding	KNN	21.06	3.43	3.34	5.51	2.62
	KF	30.38	9.14	5.81	13.24	7.25
	SVM	33.32	5.09	4.01	16.11	7.30
	KK	8.29	1.71	1.38	2.40	1.52

slightly better than the KF and SVM methods. The conclusion achieved based on Table 2 is consistent with that indicated by Fig. 12 and Table 1.

In order to further prove the significance of the proposed predictive method performing better than other three traditional methods, the F-test is introduced into the test. In the F-test, the F values can be calculated by $S^2_{l\arg er} \Big/ S^2_{smaller}$. Because the S^2 calculated using the data obtained by the KK method is always smaller than the other three methods, they are the denominators in the formula. Table 3 shows the calculated F values.

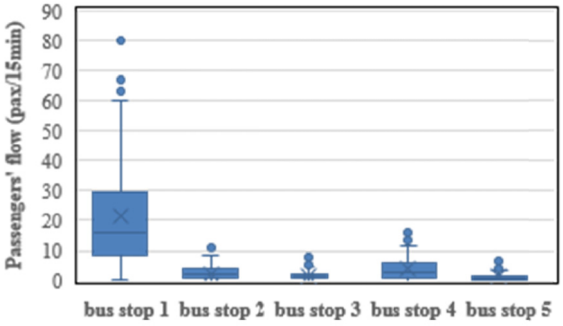
TABLE 3. F values.

Value of F		bus stop 1	bus stop 2	bus stop 3	bus stop 4	bus stop 5
Alighting	KNN/KK	4.56	4.84	4.52	4.20	4.03
	KF/KK	4.08	17.72	20.03	35.88	41.61
	SVM/KK	3.07	10.41	10.87	35.34	17.69
Boarding	KNN/KK	6.45	4.00	5.84	5.25	2.99
	KF/KK	13.43	28.45	17.70	30.38	22.84
	SVM/KK	16.16	8.82	8.44	44.96	23.18

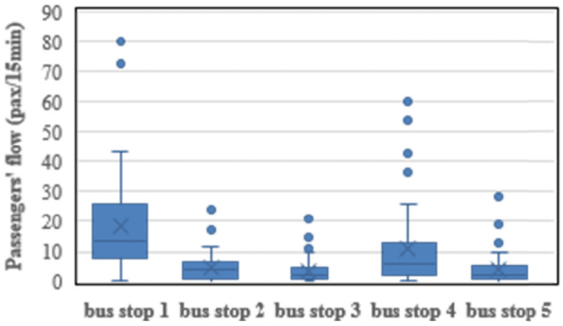
All the F values listed in Table 3 are more than 2.99, which is larger than 1.37 (according to the F Distribution Table f1 = f2 = 68). Therefore, the KK method is significantly better than the other three methods.

More detailed error distribution of the three predicted method is shown in Fig. 13, illustrating that the KK method performs stably and accurately to predict the passengers' flow under variable scenarios in terms of average absolute errors and mid-value. Therefore, the above analysis clearly shows that the proposed KK method has better performance in predicting passengers' flow in public transit.

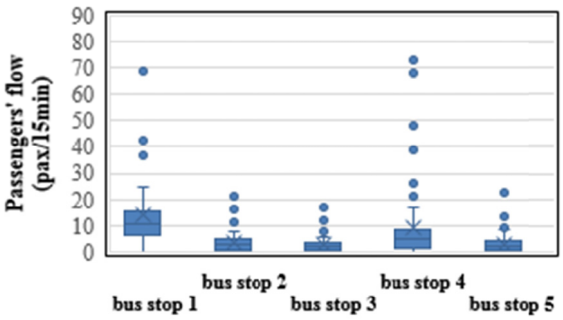
The program was run for several times. The KNN method requires less than 14 s; the KF method needs less than 4 s. Because in the proposed predictive framework, the KNN method is used once and the KF method is used twice, the efficiency of the proposed method is inevitably slightly less by approximately 20 s. However, it still can meet with



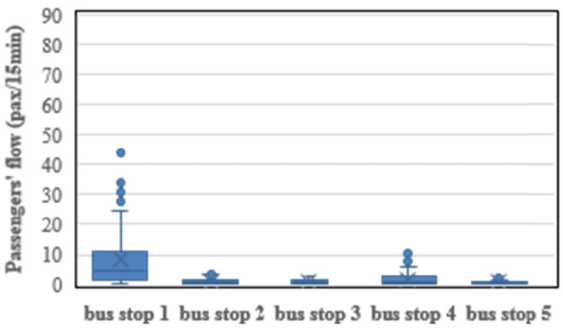
(a) KF method



(b) KNN method



(c) SVM method



(d) KK method

FIGURE 13. Comparison of distribution characteristics of absolute errors for the alighting passengers' flow at the five stops by different methods.

the requirement of real time prediction. In addition, other languages (C, JAVA, etc.) can still improve the efficiency of the method. The computer used is a laptop with Intel® Core i3-3120M CPU @2.50GHz processor and installed memory (RAM) of 4.00GB (2.32GB usable).

V. CONCLUSION

A novel concept was proposed to predict the passengers' flow of the public transit based on the historical data and data in current day. This research integrated the KNN method and KF method into a framework to make fusion of the predicted values predicted by these two methods. In the fusion framework, the weight coefficients can be calculated dynamically by the KF method again. The KF method performs badly when the data change greatly, and the KNN method cannot handle the fluctuating data very well; however, the proposed fusion model can handle these two cases based on dynamically adjusting the weight coefficient values. Based on the real data collected at the five stations on Line 3 in Changchun City, a case study was conducted by the KNN, KF, SVM, and proposed KK methods. In order to test the predicting method in different scenes, two types of stations are selected including a station located at the near railway station as a transportation junction and four normal stations for providing services for the commuting passengers. According to the test results and analysis, the proposed KK method can improve the accuracy of the predicted values for passengers' flow of both alighting passengers and boarding passengers. In addition, the KK method performs best among these four predicting methods, while the KNN is better than the KF method in predicting the values of passengers' flow. In addition, at the Changchun Station, the index of MRE is smaller than at the other four stations, while the RMSE index at the Changchun Station is much larger than at the other four stations. In addition, the hit rate, in terms of MRE by the KK method, is 92.1 within 10%, significantly larger than that obtained by the KNN, KF, and SVM methods with 72.2, 51.8 and 65.3, respectively. Therefore, the proposed KK method more accurately predicted the values of passengers' flow by dynamically obtaining the weight coefficient values. Theoretically, the fusion prediction method can predict other traffic parameters including traffic flow, travel time, and travel speed. More tests for the KK method in other scenarios will be conducted in the near future. In addition, we will explore how the accuracy of the method varies with the length of prediction time in the future work.

ACKNOWLEDGMENT

The authors confirm that there are no conflicts of interest regarding the publication of this manuscript. All data included in this study is available upon request by contact with the corresponding author.

REFERENCES

[1] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with limited traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 653–662, Aug. 2015.

[2] M. Bernas, B. Płaczek, P. Porwik, and T. Pamuła, "Segmentation of vehicle detector data for improved k-nearest neighbours-based traffic flow prediction," *IET Intell. Transp. Syst.*, vol. 9, no. 3, pp. 264–274, 2015.

[3] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 21–34, Jan. 2016.

[4] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6164–6173, 2009.

[5] A. Cheng, X. Jiang, Y. Li, C. Zhang, and H. Zhu, "Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method," *Phys. A, Statist. Mech. Appl.*, vol. 466, pp. 422–434, Jan. 2017.

[6] S. Cheng, F. Lu, P. Peng, and S. Wu, "Short-term traffic forecasting: An adaptive ST-KNN model that considers spatial heterogeneity," *Comput., Environ. Urban Syst.*, vol. 71, pp. 186–198, Sep. 2018.

[7] M. S. Grewal, *Kalman Filtering*. Berlin, Germany: Springer, 2011, pp. 705–708.

[8] F. Guo, J. W. Polak, and R. Krishnan, "Predictor fusion for short-term traffic forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 90–100, Jul. 2018.

[9] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 61–78, May 2015.

[10] R. Ke, W. Li, Z. Cui, and Y. Wang, "Two-stream multi-channel convolutional neural network (TM-CNN) for multi-lane traffic speed prediction considering traffic volume impact," 2019, *arXiv:1903.01678*. [Online]. Available: https://arxiv.org/abs/1903.01678

[11] J. Ke, H. Zheng, H. Yang, and X. M. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *J. Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.

[12] T. Kim, H. Kim, and D. J. Lovell, "Traffic flow forecasting: Overcoming memoryless property in nearest neighbor non-parametric regression," in *Proc. IEEE Intell. Transp. Syst.*, Sep. 2005, pp. 965–969.

[13] S. V. Kumar, "Traffic flow prediction using Kalman filtering technique," *Procedia Eng.*, vol. 187, pp. 582–587, Jan. 2017.

[14] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, p. 21, 2015.

[15] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, "Road traffic forecasting: Recent advances and new challenges," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 2, pp. 93–109, Apr. 2018.

[16] S. Liang, S. Zhao, C. Lu, and M. Ma, "A self-adaptive method to equalize headways: Numerical analysis and comparison," *Transp. Res. B, Methodol.*, vol. 87, pp. 33–43, May 2016.

[17] S. Liang, S. Zhao, M. Ma, and H. Liu, "Analysis of traffic conditions in urban region based on data from fixed detectors," *Discrete Dyn. Nature Soc.*, vol. 2015, Aug. 2015, Art. no. 184049.

[18] S. Liang, M. Ma, and S. He, "Multiobjective optimal formulations for bus fleet size of public transit under headway-based holding control," *J. Adv. Transp.*, vol. 2019, Jan. 2019, Art. no. 2452348.

[19] S. Liang, M. Ma, S. He, H. Zhang, and P. Yuan, "Coordinated control method to self-equalize bus headways: An analytical method," *Transportmetrica B, Transp. Dyn.*, vol. 7, no. 1, pp. 1175–1202, 2019.

[20] S. Liang and M. Ma, "Analysis of bus bunching impact on car delays at signalized intersections," *KSCE J. Civil Eng.*, vol. 23, no. 2, pp. 833–843, 2019.

[21] S. Liang, M. Ma, S. He, and H. Zhang, "The impact of bus fleet size on performance of self-equalise bus headway control method," in *Proc. Inst. Civil Eng.-Municipal Eng.*, 2018, pp. 1–29.

[22] Q. Liu, B. Wang, and Y. Zhu, "Short-term traffic speed forecasting based on attention convolutional neural network for arterials," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 11, pp. 999–1016, 2018.

[23] X. Liu, F. Wan, L. Chen, Z. Qiu, and M. R. Chen, "Research on traffic passenger volume prediction of Sanya City based on ARIMA and grey Markov models," in *Proc. Int. Conf. Pioneering Comput. Sci., Eng. Educ.* Singapore: Springer, Sep. 2018, pp. 337–349.

[24] Z. Liu, J. Guo, J. Cao, Y. Wei, and W. Huang, "A hybrid short-term traffic flow forecasting method based on neural networks combined with K-nearest neighbor," *J. Traffic Transp. Technol.*, vol. 30, no. 4, pp. 445–456, 2018. doi: 10.7307/ptt.v30i4.2651.

[25] X. Luo, D. Li, and S. Zhang, "Traffic flow prediction during the holidays based on DFT and SVR," *J. Sensors*, vol. 2019, Jan. 2019, Art. no. 6461450.

[26] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *J. Adv. Transp.*, vol. 2019, Feb. 2019, Art. no. 4145353.

[27] X. Luo, L. Niu, and S. Zhang, "An algorithm for traffic flow prediction based on improved SARIMA and GA," *KSCE J. Civil Eng.*, vol. 22, no. 10, pp. 4107–4115, 2018.

[28] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[29] M. Ma, S. Liang, H. Guo, and J. Yang, "Short-term traffic flow prediction using a self-adaptive two-dimensional forecasting method," *Adv. Mech. Eng.*, vol. 9, no. 8, pp. 1–12, 2017.

[30] M. Ma, Q. Yang, S. Liang, and Y. Wang, "A new coordinated control method on the intersection of traffic region," *Discrete Dyn. Nature Soc.*, vol. 2016, Mar. 2016, Art. no. 5985840.

[31] M. Ma, S. Liang, and Y. Qin, "A bidirectional searching strategy to improve data quality based on K-nearest neighbor approach," *Symmetry*, vol. 11, no. 6, p. 815, 2019.

[32] M. Ma and S. Liang, "An integrated control method based on the priority of ways in a freeway network," *Trans. Inst. Meas. Control*, vol. 40, no. 3, pp. 843–852, 2018.

[33] M. Ma and S. Liang, "An optimization approach for freeway network coordinated traffic control and route guidance," *PLoS ONE*, vol. 13, no. 9, 2018, Art. no. e0204255.

[34] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[35] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, 2011.

[36] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial–temporal weighted model on MapReduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246–263, Feb. 2016.

[37] H. Z. Moayedi and M. A. Masnadi-Shirazi, "Arima model for network traffic prediction and anomaly detection," in *Proc. Int. Symp. Inf. Technol.*, vol. 4, Aug. 2008, pp. 1–6.

[38] M. Moniruzzaman, H. Maoh, and W. Anderson, "Short-term prediction of border crossing time and traffic volume for commercial trucks: A case study for the ambassador bridge," *Transp. Res. C, Emerg. Technol.*, vol. 63, pp. 182–194, Feb. 2016.

[39] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3–7, Nov. 2015.

[40] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, 1984.

[41] J. Tang, F. Liu, Y. Zou, W. Zhang, and Y. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2340–2350, Sep. 2017.

[42] M. van der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.

[43] L. Vanajakshi and L. R. Rilett, "A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2004, pp. 194–199.

[44] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[45] H. Wang, L. Liu, S. Dong, Z. Qian, and H. Wei, "A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD–ARIMA framework," *Transportmetrica B, Transp. Dyn.*, vol. 4, no. 3, pp. 159–186, 2015.

[46] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.

[47] C. Xu, Z. Li, and W. Wang, "Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming," *Transport*, vol. 31, no. 3, pp. 343–358, 2016.

[48] D. W. Xu, Y. D. Wang, L. M. Jia, Y. Qin, and H. H. Dong, "Real-time road traffic state prediction based on ARIMA and Kalman filter," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 2, pp. 287–302, 2017.

[49] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "K-nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *J. Transp. Eng.*, vol. 142, no. 6, 2016, Art. no. 04016018.

[50] S. Zhao, S. Liang, H. Liu, and M. Ma, "CTM based real-time queue length estimation at signalized intersection," *Math. Problems Eng.*, vol. 2015, Aug. 2015, Art. no. 328712.

[51] J. Zhou, G. Gu, and X. Chen, "Distributed Kalman filtering over wireless sensor networks in the presence of data packet drops," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1603–1610, Jul. 2018.

[52] Y. Zou, X. Zhu, Y. Zhang, and X. Zeng, "A space–time diurnal method for short-term freeway travel time prediction," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 33–49, Jun. 2014.

SHIDONG LIANG received the Ph.D. degree from the Transportation College, Jilin University, Changchun, China, in 2017. He is currently a Lecturer with the Business School, University of Shanghai for Science and Technology, Shanghai, China. His research interests include urban signal control and public transit management and control.

MINGHUI MA received the Ph.D. degree from the Transportation College, Jilin University, Changchun, China, in 2016. She is currently a Lecturer with the College of Automobile Engineering, Shanghai University of Engineering Science, Shanghai, China. Her research interests include traffic signal control and traffic data analysis.

SHENGXUE HE received the Ph.D. degree from the Business School, University of Shanghai for Science and Technology, China, in 2008, where he is currently an Associate Professor. His research interests include traffic modeling, ITS, and the public transit system optimization.

HU ZHANG received the Ph.D. degree from the College of Transportation, Jilin University. He is currently working on his Postdoctoral Research with the Business School, University of Shanghai for Science and Technology, Shanghai, China. His research interests include public transport management and intelligent transportation systems.