Short-Term Traffic Flow Prediction of Airspace Sectors Based on Multiple Time Series Learning Mechanism

Zhaoning Zhang, Kexuan Liu(✉), Fei Lu, and Wenya Li

Civil Aviation University of China, Tianjin 300300, China
294006021@qq.com

Abstract. Firstly, by analyzing the original radar data of the aircraft in the airspace system, the historical operation information of each sector is extracted, and the traffic flow correlation between different routes of the same sector is considered. According to the characteristics of busy sector traffic flow data, a multi-dimensional data model of traffic flow with multiple related routes in the sector is constructed. Secondly, based on the data model, a traffic flow forecasting algorithm based on multi-time series machine learning is proposed. The main core idea of the algorithm is to use the time series clustering method to reduce the dimensionality of multi-dimensional traffic flow data, and then introduce the machine learning method for concurrent training. The training result obtains the optimal classifier group through competition. Finally, the multi-optimal machine learning integrated prediction method is designed to predict traffic flow. Taking the typical busy sector in China as an example, the proposed prediction method is verified. The research results show that the prediction results are better than the traditional single time series machine learning method, and the stability of the prediction results is good, which can fully reflect the dynamics and uncertainty of short-term traffic flow between sectors in each airspace, in line with the actual situation of air traffic.

Keywords: Traffic flow prediction · Learning mechanism · Airspace sectors · Flight scheduling

1 Introduction

With the rapid development of China's air transport industry, air traffic flow continues to grow, and problems such as large airspace congestion, large-scale flight delays, high-intensity control loads, and high-frequency accidents caused by the contradiction between supply and demand have become increasingly prominent. The resulting safety hazards and economic losses have seriously restricted the sustainable and healthy development of China's air transport industry. In order to ensure the safe and efficient operation of China's airspace system, it is necessary to accurately predict the short-term traffic of airspace sectors. Based on scientifically grasping the distribution trend of air traffic flow and future trends, an effective air traffic flow management strategy is formulated. It is not only the premise of implementing tactical and pre-tactical traffic

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 425–432, 2019.
https://doi.org/10.1007/978-3-030-19086-6_47

management, but also provides method support and reference basis for control sector opening and closing configuration, control seat scheduling, and air traffic situation assessment.

The traditional short-term air traffic flow forecasting method is mainly based on historical observation flow time series data, using historical average model, moving average model, autoregressive sliding model, combined forecasting model and other statistical forecasting models [1–5], and artificial intelligence such as neural networks [6–8]. The input method of the above method is single, and the calculation process is simple, but the dynamics and uncertainty of the air traffic flow during the actual operation are not fully considered. In papers [9–11], the model of aggregate traffic flow forecasting is proposed. The model takes the traffic flow distribution situation of the airspace unit as the research object, and the number of airspace units is used as the calculation dimension, which greatly reduces the calculation amount in the number of aircrafts. The model is less affected by the uncertainty factor and the prediction time range is long. However, the input parameters of the model are a set of static historical average values, and the uncertainty of the parameters is not considered. Therefore, there is a certain difference between the prediction effect and the actual operation.

It can be seen that traditional research often uses the data of the predicted route itself as the research object for analysis and modeling, and the interaction between adjacent routes is generally less studied. However, the traffic flow between multiple routes in the same area obviously has a certain correlation. Therefore, this study takes multiple routes in the same area as the research object, fully considering the dynamics of traffic flow in the actual operation of airspace system. Taking full account of the dynamic, time-varying and uncertainties of traffic flow in the actual operation of China's airspace system, this paper proposes a traffic flow prediction model based on the integrated learning concept of hybrid multi-machine learning strategy. This model considers traffic flow forecasting as multidimensional time series data mining. Through deep analysis of radar data, clear spatial sector historical traffic data and initial model parameters, according to the established multi-sector aggregate traffic flow model and parameter estimation and update method, the traffic distribution of airspace sectors in a certain time and space range in the future It is predicted to provide scientific theoretical basis and data support for the efficient management of tactical and pre-tactical air traffic flow.

2 Sector Traffic Flow Data Analysis

The experimental objects in this study selected the traffic flow data of the main segments of each route. The original information collected by each route is based on the flight, including the route number, flight signage, arrival time, and departure time. First, the number of flights per minute of each route is counted, and the traffic flow data is divided into a unified window for a certain length of time. Finally, the metrics of the traffic are separated into different levels.

Let the set of M traffic flow data in a certain sector be $Z = \{Z_1, Z_2, \ldots, Z_M\}$ and $Z_i \in Z$ is traffic flow time series information system composed of traffic flow information of the i-th route in Z is denoted as $Z_i = (W_i, K_i, C_i, F, a)$.

Where $W_i = \{w_{i1}, w_{i2}, \ldots, w_{in}\}$ represents a set of n time-ordered traffic flow timing segments in the i-th route, and each time-series data segment has a length of s time units. The data of the j-th time window of the i-th route is expressed as $w_{ij} = \left\{t_{\left(w_{ij},1\right)}, t_{\left(w_{ij},2\right)}, \ldots, t_{\left(w_{ij},s\right)}\right\}$, $w_{ij} \in W_i$, $\left|w_{ij}\right| = s$, and the predecessor relationship is satisfied between $t_{\left(w_{ij},k+1\right)}$ and $t_{\left(w_{ij},k\right)}$.

$K_i = \{k_{i1}, k_{i2} \ldots, k_{in}\}$ represents a set of n predecessor traffic flow time series data in the flow data sample of the i-th route.

$C_i = \left\{C_{i1}^h, C_{i2}^h, \ldots, C_{in}^h\right\}$ represents the set of traffic flow values in consecutive h min after each piece of historical traffic flow time series data in $K_i = \{k_{i1}, k_{i2} \ldots, k_{in}\}$, and its value is determined by the map g: $k_{ij} \to c_{ij}^h$, where j ∈ {1, 2, …, n}.

F is a mapping, F: $w_{ij} \to k_{ij}$, is a one-to-one mapping relationship between any j-th time series data k_{ij} in K_i and the j-th time series data segment w_{ij} in W_i. Can be recorded as $k_{ij} = F\left(w_{ij}\right) = \left\{f\left(t_{\left(w_{ij},1\right)}\right), f\left(t_{\left(w_{ij},2\right)}\right), \ldots, f\left(t_{\left(w_{ij},s\right)}\right)\right\}$.

The time series ***K***i set corresponding to the data sample Zi can be recorded as the traffic local time series data matrix model ***K***i as shown in the formula (1).

$$K_i = \begin{pmatrix} k_{i1} \\ k_{i2} \\ \vdots \\ k_{in} \end{pmatrix} = \begin{pmatrix} F(w_{i1}) \\ F(w_{i2}) \\ \vdots \\ F(w_{in}) \end{pmatrix} = \begin{pmatrix} f\left(t_{(w_{i1},1)}\right) & f\left(t_{(w_{i1},2)}\right) & \cdots & f\left(t_{(w_{i1},s)}\right) \\ f\left(t_{(w_{i2},1)}\right) & f\left(t_{(w_{i2},2)}\right) & \cdots & f\left(t_{(w_{i1},s)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ f\left(t_{(w_{in},1)}\right) & f\left(t_{(w_{in},2)}\right) & \cdots & f\left(t_{(w_{in},s)}\right) \end{pmatrix} \tag{1}$$

Where: each row is represented as a horizontal transpose of the s data corresponding to the k_{ij} data set.

Since the traffic flow between the routes affects each other, historical traffic data of a single route is not directly used to predict future traffic, but the route and its adjacent route history data are modeled. The traffic local time series data matrix (i.e., Z1, Z2, Z3, …, Zm) of the M routes in the Z set is combined and converted into a column of data according to the sequence of the start time of each time series. Finally, the future h min traffic flow value $C_i = \left\{C_{i1}^h, C_{i2}^h, \ldots, C_{in}^h\right\}$ of each time window of the predicted i-th route is added to the corresponding time period as a column of decision data relative to the i-th route prediction. A traffic flow prediction data set matrix model PD as shown in the Eq. (2) is obtained.

$$PD = \begin{pmatrix} k_{11} & k_{21} & \cdots & k_{M1} & C_{i1}^h \\ k_{12} & k_{22} & \cdots & k_{M2} & C_{i2}^h \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k_{1n} & k_{2n} & \cdots & k_{Mn} & C_{in}^h \end{pmatrix} \tag{2}$$

Where: j ∈ {1, 2, …, n}, i ∈ {1, 2, …, M}.

In the PD matrix, in addition to the decision column Ci, the other columns represent traffic flow records for different routes. Each row of data in the PD matrix (except for the decision column) represents traffic flow data for the same time window of M routes in the same region. Line j represents the time series data of the j-th time window of each route and the traffic flow of the i-th route in the future h min. The meanings of some of the symbols involved are shown in Table 1.

Table 1. The meaning of symbolic.

Symbolic	Meaning
Z	Collection of traffic flow data within a sector
Z_i	Traffic flow time series information system composed of traffic flow information of the i-th route
w_{in}	Time-sequential n traffic flow timing segments in the i-th route
W_i	A collection of w_{in}
k_{in}	n predecessor traffic flow time series data in the traffic data sample of the i-th route
K_i	A collection of k_{in}
C_{in}^h	Set of traffic values in consecutive h min after each piece of historical traffic flow time series data in K_i
C_i	A collection of C_{in}^h

3 Machine Learning Model

The idea of traffic flow multi-machine learning competition method is shown in Fig. 1.

Firstly, the multi-dimensional time series modeling is carried out with historical data, and the traffic flow data in the research route is converted into a multi-dimensional time series matrix model.

Then time series clustering method is used to time series clustering, and the multidimensional time series matrix model is reduced to a classic two-dimensional information table.

Then, the two-dimensional information table is imported into the multi-machine learning group for learning, and the learned knowledge is competed by the test data to generate multiple optimal classifiers. The practice data is imported into the optimal classifier and multiple prediction results are output.

Finally, the multiple prediction results are integrated and applied in practice. If the actual prediction results are not good, the training data set is updated and the optimal classifier is retrained.

3.1 Introduction to the Optimal Classifier Acquisition Algorithm

Input: Historical traffic flow data set for related M routes $Z = \{Z_1, Z_2, \ldots, Z_M\}$.

Alternative machine learning algorithm classifier set as: $L = \{L_1(), L_2(), \ldots, L_n()\}$.

Output: Optimal machine learning algorithm classifier set $L'' \subseteq L$.

Step 1 Historical traffic flow data set $Z = \{Z_1, Z_2, \ldots, Z_M\}$ of related M routes. The Z transformation is preprocessed into the multi-dimensional time series matrix model PD matrix form of traffic flow.

Step 2 Time series clustering method proposed in paper [12] is used to perform time series clustering on the above PDs. Each sub-time series k_{ij} cluster in PD is transformed into a class integer V_{ij}, so that PD is transformed into classical two-dimensional information. Formula (3), recorded as G_{PD}.

$$PD = \begin{pmatrix} k_{11} & k_{21} & \cdots & k_{M1} & C_{i1}^{h} \\ k_{12} & k_{22} & \cdots & k_{M2} & C_{i2}^{h} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k_{1n} & k_{2n} & \cdots & k_{Mn} & C_{in}^{h} \end{pmatrix} \xrightarrow{\text{Clustering}} G_{PD} = \begin{pmatrix} V_{11} & V_{21} & \cdots & V_{M1} & C_{i1}^{h} \\ V_{12} & V_{22} & \cdots & V_{M2} & C_{i2}^{h} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ V_{1n} & V_{2n} & \cdots & V_{Mn} & C_{in}^{h} \end{pmatrix} \tag{3}$$

Finally, the G_{PD} is divided into a test data set G'_{PD} and a training data set G''_{PD}.
Step 3 introduces G'_{PD} into the alternative classical machine algorithm set $L = \{L_1(), L_2(), \ldots, L_n()\}$ for machine learning. After learning $L' = \{L_1()', L_2()' \ldots, L_n()'\}$.
Step 4 introduces G''_{PD} into the alternative classical machine algorithm set $L' = \{L_1()', L_2()' \ldots, L_n()'\}$. And the statistical accuracy of each algorithm is recorded as $\tau_{L'_1}, \tau_{L'_2}, \ldots, \tau_{L'_n}$. And the threshold ω is set, if $\tau_{L'_i} > \omega$, i ∈ {1, 2, …, m}, Then the algorithm $\tau_{L'_i}$ is reserved as the optimal algorithm set L'' member. $L'' \subseteq L$.

3.2 Introduction to the Integrated Prediction Algorithm in This Prediction Model

Input: The real-time traffic flow data flow of the relevant M route x time window is $k_{1x}, k_{2x} \ldots, k_{Mx}$.

Alternative machine learning algorithm classifier set L''. Note: Let the number of algorithms in A lg″ be set to n, n < m. and re-mark the candidate algorithm set as: $L'' = \{L'_1(), L'_2(), \ldots, L'_n()\}$, the prediction accuracy is respectively recorded as $\{\tau_1, \tau_2, \ldots, \tau_n\}$.

Output: the future h min traffic flow value β of the x time window of the i-th route.

Step 1 For real-time traffic flow data: perform clustering category matching [13], convert the time series string $k_{1x}, k_{2x} \ldots, k_{Mx}$ into an integer array V1x V2x ⋯ VMx. Record as array Vx = {V1x V2x … VMx}.
Step 2 imports the array Vx into $L_1()', L_2()', \ldots, L_n()'$. The prediction results are respectively recorded as $L_1(V_{1x})', L_2(V_{2x})', \ldots, L_n(V_{nx})'$. Where $L_i(V_x)'$ represents the prediction result of the i-th optimal prediction algorithm for real-time traffic flow data V_x.
Step 3 The prediction result is $\beta = \left| \sum_{i=1}^{n} \tau_i L_i / \sum_{i=1}^{n} \tau_i \right|$, where τ_i is the algorithmic weight of L_i.

4 Experiment and Result Analysis

4.1 Experimental Settings

Alternative machine learning algorithms include: proximity algorithm, Bayesian algorithm, neural network, support vector machine, etc. ω is the threshold parameter for predictive evaluation, ω = 0.8. Select a busy high-altitude sector as the research object, and use the radar data from January 16 to February 16, 2018 to establish a database to predict the traffic flow from 8:00 to 13:00 on February 17. The algorithm of [6] compares the results, the result is the comparison of the actual traffic level and the predicted traffic level.

4.2 Experimental Results

The results of the I and II route experiments are shown in Figs. 1 and 2, where the horizontal axis represents time, at 10 min intervals, and the vertical axis represents traffic flow levels. Every 10 min, there is a corresponding actual route traffic flow, single dimension. A time-series predicted traffic flow and a comparison of traffic flows predicted by the proposed method. Among them, SIMPLE and MULTIBLE are the results of single machine and multidimensional time series hybrid machine learning model prediction. ACTUAL represents real traffic flow.

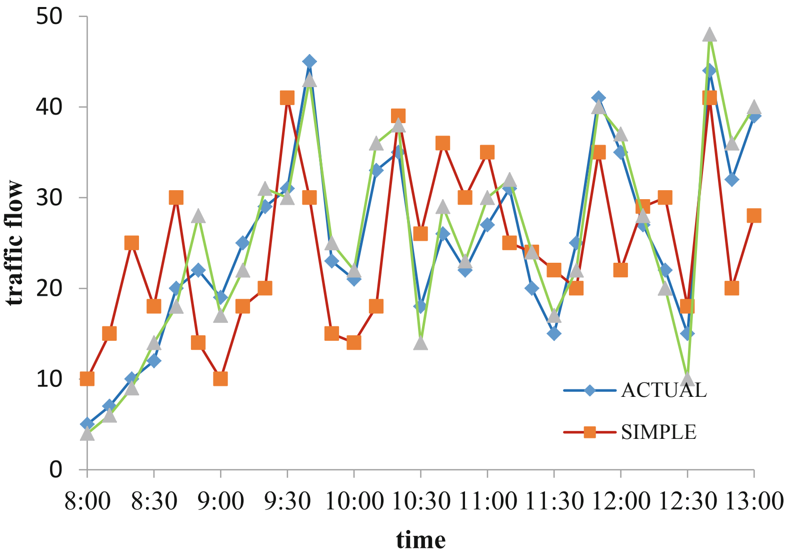

Fig. 1. Comparison of effectiveness between actual flow and two methods for route I.

As shown in Figs. 1 and 2, for the I and II routes, the accuracy predicted by the single time series method under the selected time period is 70.49% and 59.98%, respectively, and the multi-dimensional time series hybrid machine learning method is used to predict the results. The accuracy rates are 87.01% and 88.94%, respectively.

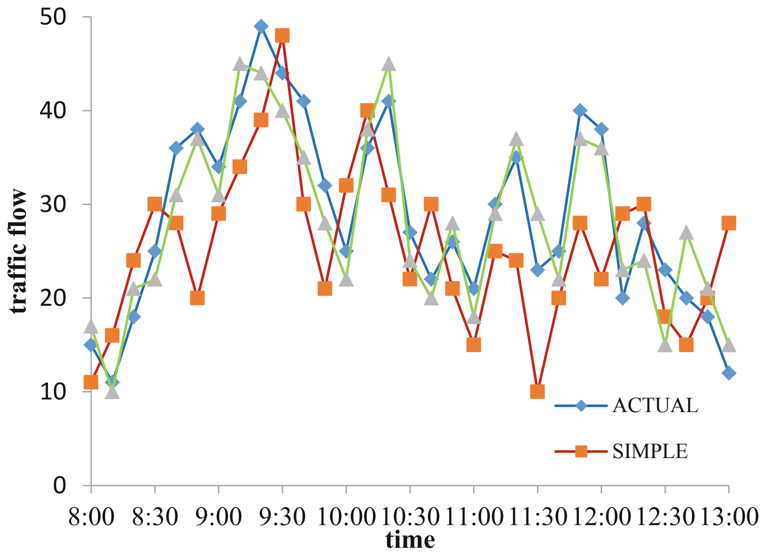

Fig. 2. Comparison of effectiveness between actual flow and two methods for route II.

The algorithm in this paper is closer to the actual traffic flow of the predicted route, and its fitting degree is higher, and the fitting degree of the flow curve of the comparison algorithm is far away. This shows that the proposed algorithm has obvious advantages. Compared with the multi-dimensional time series, the single time series algorithm often cannot obtain enough relevant route information. Obviously, only all relevant route information in the comprehensive study area can carry out more effective knowledge mining. The multi-dimensional time series clustering method proposed in this paper. It can guarantee the retention of relevant information in different dimensions, and can effectively reduce the data dimension and utilize a variety of classical machine learning methods.

5 Conclusions

Aiming at the data characteristics of route traffic flow prediction, this paper proposes a multi-dimensional time series data mining traffic flow prediction model. This model can comprehensively consider the correlation of the relevant routes of the same sector, and adopts integrated learning ideas to compete through multi-machine learning. It greatly improves the accuracy of traffic flow prediction. Through prediction practice, it is proved that the prediction effect of this model is better than the traditional single time series machine learning algorithm.

Acknowledgements. The authors were supported by the National Natural Science Foundation of China (No. 71701202).

References

1. Yu, B., Wu, S., Wang, M., et al.: K-nearest neighbor model of short-term traffic flow forecast. J. Traffic Transp. Eng. **12**(2), 109–115 (2012)
2. Sun, X., Liu, T.: Short-term traffic flow forecasting based on a hybrid neural network model and SARIMA model. J. Transp. Syst. Eng. **8**(5), 32–37 (2008)
3. Han, W., Wang, J.-F., Gao, Y.-G., et al.: Forecasting and analysis of regional traffic flow in space and time. J. Highw. Transp. Res. Dev. **24**(6), 92–96 (2007)
4. Chang, G., Zhang, Y., Yao, Y.-d.: Short-term traffic flow forecasting model for regional road network based on spatial-temporal depend. J. Tsinghua Univ. (Sci. Technol.) (2), 215–221 (2013)
5. Xie, J., Wu, W., Yang, X.: A PDL model used for short-term traffic flow forecasting. J. Tongji Univ. (Nat. Sci.) **39**(9), 1297–1302 (2011)
6. Wang, Y.-G., Li, H.: Prediction model of civil aviation incidents based on grey neural network. China Saf. Sci. J. **22**(3), 10–15 (2012)
7. Ou, T., Zhou, C.: Situation assessment model of civil aviation safety based on neural network and its simulation. J. Saf. Sci. Technol. **7**(2), 34–41 (2011)
8. Li, Y.-b., Li, C., Song, X.-h.: Prediction model of improved artificial neural network and its application. J. Cent. S. Univ. (Sci. Technol.) (5), 1054–1058 (2008)
9. Kumar, S.V.: Traffic flow prediction using Kalman filtering technique. Procedia Eng., 187 (2017)
10. Wu, Y., Tan, H., Qin, L., Ran, B., Jiang, Z.: A hybrid deep learning based traffic flow prediction method and its understanding. Transp. Res. Part C **90** (2018)
11. Ratrout, N.T.: Short-term traffic flow prediction using group method data handling (GMDH)-based abductive networks. Arab. J. Sci. Eng. **39**(2) (2014)