A multi-step airport delay prediction model based on spatial-temporal correlation and auxiliary features

Hao Zhang1 | Chunyue Song1 | Jie Zhang2 | Hui Wang1 | Jinlong Guo1

1 State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, People’s Republic of China
2 Polytechnic Institute, Zhejiang University, Hangzhou, People’s Republic of China

Abstract
Airport delay prediction system is fundamental to intelligent air traffic management. However, the prediction of airport delay is affected strongly by spatial-temporal dependencies and other exogenous dependencies, which would bring serious challenges in prediction. In this paper, the APR-LSTM model is proposed to address these challenges. The model is driven by the real-world flight operation data. In the proposed prediction model, the dynamic air traffic delay networks are fed into PageRank algorithm to capture the dynamic spatial dependencies. The input sequences of airport delay vector is weighted by spatial dependencies before sending into the long short-term memory network (LSTM) and sequence to sequence model, which can realise the multi-step prediction and the joint mining of spatial-temporal dependencies. Subsequently, the temporal attention mechanism and auxiliary features are introduced to obtain the long-term temporal dependencies by exploring the relevance of different time steps and improve the accuracy of prediction model. Furthermore, the results of comparative experiments indicated that the proposed model achieve the best performance compared with other benchmark models. Finally, the spatial-temporal correlation of different airports, the delay characterisations of networks and the application of prediction results are demonstrated in detail, enabling the interpretability of APR-LSTM model.

1 | INTRODUCTION

Due to the convenience, the air transportation industry has experienced a rapid growth and the demands have continued rising in recent years. In 2018, Chinese air transportation completed serving 611.73 million passengers [1]. However, airport infrastructures and airspaces are unable to satisfy the growing demands, leading to severing flight delay and delay propagations. Flight delay causes increasingly serious conflicts between passengers and airline companies, resulting in economic and resource losses [2] and limiting the development of Chinese air transportation industry. In order to mitigate the negative impacts of delay, the delay prediction system is developed and becomes the key of intelligent air traffic management (IATM). The goal of prediction system is to improve the robustness of air transportation system which could help make decisions in advance to optimise flights, airports and airspace resources.

The air transportation system is complex and random, containing multi parts such as flights, airports, air routes and airspaces. These parts along with several dynamic environmental factors can affect airport delay. The research found that Chinese air traffic network is a small-world and scale-free. The connection between Chinese domestic airports is even tighter than other countries, that means delay will propagate quickly among the entire network [3].

Besides, the airport delay has strong spatial-temporal dependencies. For spatial dependencies, one airport delay is influenced by the surrounding airports where there are flight routes, which make them not exist in isolation. For temporal dependencies, the airport delay is related to temporal factors, which would change by time, like holidays, weekends, airflow and weather. Therefore, the dynamic, complex and strong spatial-temporal correlation characteristics of air delay will bring huge challenges to prediction.
Many researchers have focused on airport delay prediction. Data is essential and fundamental for verifying and training model. Since Chinese flight dataset is not publicly available, there is a lack of research on Chinese aviation system. Data used for current researches can be obtained from the Federal Aviation Administration and the Bureau of Transportation Statistics. The current airport delay prediction models can be divided into two categories: one is statistical analysis mechanism models and the other is data-driven models such as machine learning and deep learning models.

The common statistical mechanism prediction models consist of Markov Chain, Bayesian network, queuing and network decomposition model and virus propagation model such as SIS models. These models are interpretable and focus on the process of delay propagation. However, these models, whose objects are mainly single hub airport, cannot handle the high-dimensional large dataset. Besides, it is difficult to get the multi-step delay prediction of each airport in the entire network.

With the development of data mining technology, machine learning and deep learning models become popular in airport delay prediction field. The machine learning models used for prediction consist of random forest, support vector machine(SVM) and GBDT. As for these models, auxiliary features such as weather and airport attributes can be taken into the models easily. However, few studies have taken the influence of other airports into account. Besides, those models will produce the rolling errors in multi-step predictions.

Deep learning models have higher prediction accuracies than machine learning models. The deep learning prediction models used for prediction include convolutional neural network (CNN), recurrent neural network (RNN), long short term memory network (LSTM) and gate recurrent unit (GRU). In addition, sequence to sequence (Seq2Seq) model can be designed to combine with various deep learning models to avoid the accumulation of errors caused by rolling prediction.

Airport delay prediction is usually considered as time-series prediction problems, ignoring the spatial dependencies of other airports in the network. Airport delay has strong dynamic spatial-temporal dependencies so that the spatial dependencies can help improve the prediction model accuracy, especially for long-term traffic status prediction. Many researchers have explored the spatial correlation by setting the model structure reasonably. The common method is utilising the correlation coefficients to analyse the correlation between airports and obtain airport sets with strong spatial correlation, which can only measure the correlation in isolation and the spatial correlation is static. Other methods, such as the HITS and community detection algorithms, including Girvan Newman and random walk, are utilising the complex network theory, to capturing spatial dependencies. Those models can identify the important sections and analyse their dynamic spatial characteristics in the network. Recently, many researches apply the deep learning model such as CNN to capture complex spatial-temporal correlations. However, the air transportation network with non-Euclidean topology graph structure is not suitable for using CNN. Graph convolution network (GCN) can avoid the insufficiencies of CNN but is not capable of accommodating the air transportation network with dynamic adjacency matrix.

Though many researchers focus on the airport delay prediction field and get great academic achievements, there are still some deficiencies in this field:

1. The previous prediction model, whose objects are usually a single airport and the lag is one-step, does not have guiding significance for air traffic controllers. Predicting the trend of multi-airport delay can offer more alert information and is more useful in the practical IATM.

2. Capturing the dynamic spatial-temporal correlation simultaneously is so challenging that deserves more comprehensive research. The current research generally regards airport delay prediction as time-series prediction problems, ignoring the spatial dependencies of other airports in the network. However, airport delay has strong dynamic spatial-temporal dependency. The core of these prediction problems is how to capture the spatial correlations and temporal evolution rules simultaneously.

3. Data-driven models have higher prediction accuracy but poor interpretability, making it difficult to get the impacts of different model components and the important segments of topological network.

4. China’s traffic network has its own characteristics, which is a small-world and scale-free. That means the connections between Chinese airports are even tighter than other countries. Thus, the proposed model is suitable for the characteristics of Chinese airports delay states. Besides, note that in this paper, the Chinese airports only refers to Chinese mainland airports.

In order to improve the shortcomings of the current research, we propose a deep learning prediction model named PageRank LSTM-Seq2Seq attention (APR-LSTM). This model can acquire the dynamic spatial correlations and the temporal evolution rules simultaneously, as well as taking the auxiliary features into consideration. The proposed model includes three parts: spatial correlation extraction module, multi-step delay prediction module and temporal attention mechanism module. The spatial correlation extraction module receives the dynamic adjacency matrix and gets the spatial dependencies through PageRank. The multi-step prediction module, which is utilised to realise multi-step delay prediction, consists of LSTM and Sequence-to-Sequence model. Thirdly, the temporal attention mechanism is integrated into model to explore the relevance of different time steps. Auxiliary features are sent into this model to improve accuracy. The comparative experiments result illustrate that the proposed model has the best performances compared with other models. Finally, the impact of different model’s part and the application of prediction results are demonstrated in detail. The main contributions of this paper are summarised as follows.
1. The air traffic delay networks are modelled as weighted directed graphs based on the real-world domestic flight operation data, and PageRank are utilised to explore the dynamic spatial correlation, improving the limitation of traditional methods.

2. Based on the spatial features caught by PageRank, the input sequences are weighted before sending into the LSTM-Seq2Seq model, which implements the joint mining of the dynamic spatial-temporal dependencies simultaneously and realise multi-step airports delays prediction.

3. The temporal attention mechanism and one-hot encoding auxiliary features are integrated into the model and the importance of each time step in historical data can be obtained.

4. The proposed model has significant superiority in model accuracy. The impacts of different model’s parts and the application of prediction results are revealed in detail, enabling the deep learning model more interpretable.

5. The proposed model conforms to the characteristics of Chinese airports delay states.

The main framework of this paper is organised as follows. Chapter 2 gives variables and prediction problem definitions. Chapter 3 introduces the multi-step airport delay prediction model in detail. Chapter 4 elaborates the dataset we used and conducts the comparative experiments to verify the superiority of proposed model. The impacts of different model’s parts and the application of prediction results are demonstrated in detail. Chapter 5 summarises the paper and offers the directions of future research.

2 | PRELIMINARY

In this section, the definitions of important variables and the prediction problem of multi-step airport delay are introduced.

2.1 | Variable definitions

2.1.1 | Flight delay

In China, the flight delay is defined as the real departure time \( r_{dep\_time} \) minus the scheduled departure time \( p_{dep\_time} \) if the difference is larger than 15 min, otherwise, it would be defined as 0. Flight delay can be calculated as

\[
dep\_delay = \max (r_{dep\_time} - p_{dep\_time} - 15, 0) \quad (1)
\]

2.1.2 | Airport delay

The average delay of airports at each hour are employed to analyse and predict in this paper. The airport \( v_i \) delay at the time \( t \) is defined as the average delay of all departure flights’ delay during 1 h interval, denoted as \( x_i(t) \). \( N \) is the number of airports in the network. The delays of the whole network at the time \( t \) is defined as the vector of \( x(t) = [x_1(t), x_2(t), ..., x_N(t)]^T \in \mathbb{R}^N \)

2.1.3 | Air traffic delay network

The air traffic delay networks can be modelled as directed weighted graphs \( G = (V, E, W) \). The graphs contain three parts: nodes, edges, and weights. The node set \( V = \{v_1, v_2, ..., v_N\} \) represents airports and \( N \) is the number of airports in the network. The route set \( E = \{(v_i, v_j)|e_1, e_2, ..., e_M\} \) represents air routes and \( M \) is the number of routes. \((v_i, v_j)\) denotes an origin-destination pair (OD pair). If airport \( v_i \) and airport \( v_j \) have route, the weight at time \( t \) is the average delay of the route \((v_i, v_j)\), denoted as \( a_{i,j,t} \). As the adjacency matrices are time-varying, \( A_t \) represents the status of air traffic network at the time \( t \) and the definition is as follow:

\[
A_t = \begin{cases}
    a_{i,j,t} & (v_i, v_j) \in E \\
    0 & (v_i, v_j) \notin E
\end{cases} \quad (2)
\]

The definition of the adjacency matrix at all times is \( W = \{A_t\} \in \mathbb{R}^{N \times N \times T} \) and \( T \) is the length of time.

2.1.4 | Auxiliary feature

Airport delay is related to time of day, season, weather and the scales of airport. Auxiliary features can improve the prediction accuracy [25]. The categories of auxiliary features taken into the model are shown as follows.

- Month: Chinese air transportation system has two flight schedules and the most of serious delays occur in Jun, July, August and September. The season can be divided into two categories, the peak months are from June to October and other months are off-peak month [1].
- Time of day: Airport delay increase since 6 am and peak around 20:00. The 24 h can be intuitively divided into two periods: the 18:00 to 00:00 are busy-hour and others are off-busy-hours.
- Day of week: Delay is severe on Monday, Tuesday and Friday and the most of serious delays occur in Jun, July, August and September.
- Weather: Weather is essential for flight operation. Bad weather features should be mined in the model to indicate the weather status. The XGBoost algorithm is utilised to analyse the importance of different auxiliary features. The results indicate that the weather features affect airport delays seriously, including fog, thunderstorm, heavy rain, flying sand and snow, which can be taken into the prediction model.
- Whether is holiday: Holiday leads to a sharp increase of passengers and exacerbates airport delays.
- The scale of airport: According to the annual throughput of airport, airport is divided into large, medium and small airports.
2.2 Problem definition

Airport delay is not only related to the trend of airport delay in the previous period, but also has spatial dependencies with other airports in the network. Based on previous research, the task of multi-step airport delay prediction model use past airport delay to predict each airport delay in the future period. The input sequences of models are the current and previous sequences of airport delay and the dynamic spatial correlation. The outputs are the trend of all airports delays in the future period.

Let \( x^*_p(t) \) denotes the predicted delay value of airport \( v_i \) at the time \( t \). \( x^*(t) \) represents the predicted delay status of the network at the time \( t \). Assume the size of input step is \( h \) and the size of output step is \( p \). The definitions of input sequence, the prediction output sequences are show in Equation (3) and (4).

\[
X_h = [x(t-h+1), x(t-h+2), \ldots, x(t)] \in \mathbb{R}^{N \times h}
\]  

\[
X^*_p = [x^*(t+1), x^*(t+1), \ldots, x^*(t+p)] \in \mathbb{R}^{N \times p}
\]  

Airport delay is the typical spatial-temporal data. The dynamic spatial correlation should be taken into consideration to improve the accuracy of model. \( W \) is the set of adjacency matrices at all the times. The input sequences of the adjacency matrix \( A_h \) is defined in Equation (5).

\[
A_h = (A_{t-h+1}, A_{t-h+2}, \ldots, A_t) \in \mathbb{R}^{N \times N \times h}
\]  

\( F_p \) represents the set of auxiliary features in the period needed to predict, which is defined in Equation (6).

\[
F_p = (F(t+1), F(t+2), \ldots, F(t+p))
\]  

In summary, the inputs of the model are \( X_h, A_h \) and \( F_p \), the output predicted sequence is \( X^*_p \). The goal of prediction model is to find the optimal function \( g(\cdot) \) to minimise the error between the actual output sequences and the predicted sequences \( X^*_p \).

The prediction model is defined in Eq. (7).

\[
X^*_p = g(X_h, A_h, F_p)
\]  

3 MULTI-STEP AIRPORT DELAY PREDICTION MODEL

To overcome the disadvantages of aforementioned research, A multi-step airport delay prediction model named as PageRank LSTM-Seq2Seq Attention (APR-LSTM) is proposed. The model includes three parts: the spatial correlation extraction module, the multi-step prediction module and the temporal attention mechanism module.

Firstly, the dynamic adjacency matrix is employed via PageRank to capture the dynamic spatial correlation. Based on the spatial features caught by PageRank, the input sequences are weighted and sent into the LSTM-Seq2Seq model. Secondly, the temporal attention mechanism is utilised to calculate the weighted hidden state vectors. Finally, the weighted hidden state vectors, the hidden state vectors and auxiliary features are sent into the fully connected layer to obtain the multi-step delay prediction sequences. The structure of model and the is shown in Figure 1, and the principles of different parts are introduced in detail as follows.

3.1 Spatial correlation extraction module

The air traffic networks are much more intricate and its adjacency matrices are time-varying, leading to much more difficult in capturing the spatial correlations. The purpose of exploring the spatial correlation is to clarify the impact of each airport on the network and identify the important sections of the network. The limitations of traditional methods have been illustrated in Chapter 1. In order to solve the deficiency of traditional methods, PageRank algorithm is applied to obtain the dynamic spatial correlation of each airport [26].

In Google, PageRank was used to measure the importance of web pages in the search engine because of its ability to evaluate the importance of all nodes in a directed graph. Besides, in recently research, PageRank algorithm has been successfully used to obtain the network status and the importance of each node [27]. As we can see, the topologies of air traffic delay network is similar to websites network, the important ones are visited frequently and the delay states of them have a huge impact on the entire network. Therefore, using PageRank could explore the dynamic spatial correlation by catching the important airports.

In PageRank algorithm, \( PR(V^i) = [pr^i_1, pr^i_2, \ldots, pr^i_N] \) denotes the delay weight vector of each airport at time \( t \). \( pr^i_t \) is the influence of airport \( v_i \) on the network delay status at time \( t \). Before iteration, each node is assigned an initial weight value.
$pr^*_t$, which usually is equal to $1/N$. $d$ is the damping coefficient and the probability of clicking new node is $1 - d$. The iterative equation of $pr^*_t$ is defined in Equation (8) and (9).

$$pr^*_t = 1 - d + \frac{d}{N} \sum_{j=1}^{m} \frac{a_{j,i,t}}{S_{\text{out}}(v_j)_t} pr^*_j$$

(8)

$$S_{\text{out}}(v_j)_t = \sum_{v_i \in V_{\text{out}}(v_j)} a_{j,i,t}$$

(9)

$S_{\text{out}}(v_j)_t$ is the out-strength of the node $v_j$. $V_{\text{out}}(v_j)_t$ is the set of nodes that have edges to the node $v_j$. $m$ is the number of nodes which has the in-links of node $v_j$. The importance vector of $PR(V)_t$ is obtained until the importance of all node converge. The larger value of $pr^*_t$, the greater influence on the other airport delay status. The range of $pr^*_t$ is [0,1] and the sum of $pr^*_t$ is equal to 1. PageRank is a Markov transfer process in essential. Therefore, the results of transfer could be by iterating the state transition matrix continuously until error less than the tolerance value, which is set as $10^{-6}$. The state transition matrix $M$, at time $t$ is defined in Equation (10).

$$M_{i,j,t} = \begin{cases} 
\frac{a_{i,j,t}}{S_{\text{out}}(v_i)_t} & \text{if node } v_i \text{ and } v_j \text{ have edge} \\
0 & \text{others}
\end{cases}$$

(10)

The PageRank algorithm can mine the airports with small degree and large delays in the network, which involves the static structure of network and the delay status of routes simultaneously. In this paper, the inputs of PageRank model are the adjacency matrices $A_t$, converting the dimension from $N \times N$ to an $N \times 1$ vector. PageRank algorithm can measure the dynamic spatial propagation characteristics of delay at each time.

This paper [29] proposes the spatial attention mechanism which can determine the spatial importance of input sequences in each step to capture the spatial correlation of input sequences. $a_i^t$ is the normalisation spatial value of airport $v_i$ calculated by the spatial attention mechanism, whose range is [0,1]. The sequences of airport delay are multiplied by the spatial attention mechanism features and then send the weighted sequences into LSTM-Seq2Seq model.
It is easy to find that \( p_r \) and \( d_i \) (see [29]) represent the same meaning. But in the calculation process of \( p_r \), spatial correlation of airport delay sequences and the topology of network are taken into consideration simultaneously. PageRank has strong ability of capturing spatial correlation. Therefore, the input sequences of prediction model can be weighted by the spatial vectors calculated by PageRank algorithm. The equation of weighted input \( x'_i(t) \) is defined in Equation (11).

\[
x'_i(t) = p'_r x_i(t)
\]

(11)

\( X'_t \) is defined as the airport delay weighted sequence and \( X'_b \) represents the weighted input sequence.

\[
X'_t = (p'_r x'_1(t), p'_r x'_2(t), \ldots, p'_r x'_N(t)) \in \mathbb{R}^N
\]

(12)

\[
X'_b = (x'_1(t-b+1), x'_2(t-b+2), \ldots, x'_N(t)) \in \mathbb{R}^{N \times b}
\]

(13)

In this paper, the weighted input sequences contain the spatial-temporal correlations. The weighted input sequences are sent into LSTM and Seq2Seq model, which implement the joint mining of the spatial-temporal correlation and predict the trend of airports delay.

### 3.2 Multi-step prediction module

This part consists of LSTM-Seq2Seq. LSTM is widely used in the time series prediction problems to solve the problem while RNN cannot handle the long-distance dependence. The LSTM model adds a memory unit to store long-term status. There are three gates in LSTM model: forget gate, input gate and output gate. The structure of the LSTM model is shown in Figure 2.

The forget gate selectively discards some information on the long-term state \( C_{t-1} \). In order to get the long-term state \( C_t \), the input gate is utilised to selectively extract the information on the current state \( C'_t \). The long-term state \( C_t \) updates the state of hidden layer \( h_t \) through the output gate. \( \sigma(x) \) and tanh(\( x \)) are activation functions. Above all, the equations of LSTM at the time \( t \) are as follows.

\[
f_t = \sigma(W_f \cdot [h_{t-1}, x'_t] + b_f)
\]

(14)

\[
i_t = \sigma(W_i \cdot [h_{t-1}, x'_t] + b_i)
\]

(15)

\[
C'_t = \tanh(W_c \cdot [h_{t-1}, x'_t] + b_c)
\]

(16)

\[
C_t = f \otimes C_{t-1} + i \otimes C'_t
\]

(17)

\[
o_t = \sigma(W_o \cdot [h_{t-1}, x'_t] + b_o)
\]

(18)

\[
b_t = o_t \otimes \tanh(C'_t)
\]

(19)

\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]

(20)

\[
\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}
\]

(21)

Seq2Seq model can avoid the accumulation of errors in traditional rolling multi-step prediction models, which is mainly used when the lengths of input and output sequences are different. The model shows great performance in machine translation and multi-step prediction in transportation. [16] Thus, to overcome the challenges of multi-step prediction accumulated error, the model has been used in this article.

Seq2Seq model [30] consists of the encoder part and decoder part. The encoder part receives the input sequence \( X'_b \) and encodes it into the context vector \( C \). The decoder part gets the prediction sequence based on \( C \) and the decoder input sequence. The objective functions of the Seq2Seq prediction model is shown in Equation (22).

\[
x^* (t+i) = \arg \max \sum_{i=1}^{p} \log P(x^* (t+i)|x^* (t+1), \ldots, x^* (t+i - 1), C)
\]

(22)

The value of \( x^* (t+i) \) is dependent on the previous sequence and the vector \( C \). The combination of Seq2Seq and LSTM can reduce the accumulation of errors caused by rolling prediction model. When the length of encoder is too long, the vector \( C \) cannot store too much information so the longer-term state will be ignored, resulting in the decrease of prediction accuracy. Therefore, the attention mechanism can be introduced for long-term prediction and learn the importance of long-term data.
3.3 Temporal attention mechanism

The temporal attention mechanism can be introduced for learning the importance of long-term data and improve the interpretability of model. The concept of attention mechanism is measuring the relevance between different hidden states of encoder and decoder parts. There are some basic forms of attention mechanisms [31]. At the time $t + a$, the temporal attention mechanism are shown in Equation (23–25):

$$ e_{t+\alpha}^b = h_t^b h_{t+b} $$

$$ \alpha_{t+\alpha}^b = \frac{|e_{t+\alpha}^b|}{\sum_{b'=1}^{b-1} |e_{t+a}^b|} $$

$$ h_{t+\alpha}^b = \sum_{b'=0}^{b-1} \alpha_{t+\alpha}^{b'} h_{t+b'} $$

$e_{t+\alpha}^b$ is the similarity between $h_{t+b}$ and $h_{t+a}$, $a \in \{1, 2, \ldots, \beta\}$ and $b \in \{1, 2, \ldots, b - 1\}$. The Softmax function is used to normalise the importance score $e_{t+\alpha}^b$, $h_{t+\alpha}^b$ is the weighted output hidden state calculated by attention mechanism.

In the end of model, the weighted hidden state, the origin hidden state and the auxiliary features are sent into the fully connected layer simultaneously to obtain the final prediction sequences of airport delay.

4 EXPERIMENTS

4.1 Data description and preprocessing

The dataset used in this paper is the domestic flight operation data from Variflight company, which is not publicly accessible, contains the features of the departure/arrival airport, the planned departure/arrival time, the real departure/arrival time and flight status during 1st Jan and 31st Dec in 2016.

Abnormal and missing values can affect the prediction model’s accuracy. It’s essential to process the dataset. The data that have missing values of departure/arrival airports or time are removed and the flight delay more than 6 h is also deleted.

The processed dataset consists of 3,123,947 flight operation data, which contains 112 airports and 4680 routes and this paper mainly focuses on the airport delay. The average delays of airports and routes are calculated by the 1-h sampling interval. Air traffic delay networks are generated in the past 366 days at 1-h intervals, bringing a total of $366 \times 24 = 8784$ networks. The topology of air traffic delay network in this paper is shown in Figure 3.

Due to the difference of magnitude and distribution of airports, it is necessary to normalise the data. The max–min
normalisation method is used to scale the data at [0,1]. The input and output sequences are generated via sliding time window. The 70% of the data is used as the training dataset, the 10% of the data is used as the valid dataset and the 20% of the data is used as the test dataset in order.

4.2 Experimental settings

4.2.1 Evaluation metric

The accuracy of prediction model is measured by Equation (26) and (27): mean absolute error (MAE) and root mean square error (RMSE). The equations are as follows. \( x^*_{ij}(t) \) is the prediction value of airport \( v_i \) at the time \( t \). \( T_{test} \) is the size of test dataset.

\[
MAE = \frac{1}{NT_{test}} \sum_{t=1}^{T_{test}} \sum_{i=1}^{N} \left| x_i(t) - x^*_{ij}(t) \right| \tag{26}
\]

\[
RMSE = \sqrt{\frac{1}{NT_{test}} \sum_{t=1}^{T_{test}} \sum_{i=1}^{N} \left( x_i(t) - x^*_{ij}(t) \right)^2} \tag{27}
\]

4.2.2 Environment

The experimental system is win10 64-bit. The python is 3.7.0 and the keras version is 2.0.8. The back propagation algorithm (BPTT) is used to train network and the algorithm optimiser is Adam, which has obvious advantages in solving non-convex optimisation problems and has faster convergence speed.

4.2.3 Hyperparameters

The hidden layer dimension, batch size and input step size are adjusted on the validation set. The dimension of hidden layer affects the complexity and accuracy of model. The model with low hidden layer dimension will deteriorate the network learning ability and is easy to increase the overfitting probability when the dimension is large. The model with large input step will increase the training time greatly and the gradient disappearance is easy to occur. Too small size of input step will fail to meet the requirement of learning the temporal correlation of data. The MAE curves under different hidden layer dimensions and different input step are plotted, and it is obvious to find that the optimal dimension of hidden layer is 32, the batch size is 32 and the optimal input step \( b \) is 7.

4.2.4 Baselines

We compare the proposed model with other prediction models:

Table 2 demonstrates the MAE and RMSE of different models at different prediction steps. The phenomena can be observed from the experiment results. Compared with other comparison models, the APR-LSTM model has the best accuracy and stability at each prediction step. It’s proved indirectly that the model can capture the spatial-temporal correlation greatly.

| Model       | Metric (min) | 1 h | 2 h | 3 h | 4 h | 5 h | 6 h | Mean |
|-------------|--------------|-----|-----|-----|-----|-----|-----|------|
| ARIMA       | MAE          | 5.40| 6.02| 6.29| 6.69| 6.85| 6.98| 6.37 |
|             | RMSE         | 5.40| 6.02| 6.29| 6.69| 6.85| 6.98| 6.37 |
| XGBoost     | MAE          | 4.54| 4.64| 5.01| 5.47| 5.80| 5.96| 5.24 |
|             | RMSE         | 10.18| 10.50| 10.84| 11.15| 11.53| 12.00| 11.03 |
| LSTM        | MAE          | 4.00| 4.14| 4.24| 4.32| 4.45| 4.50| 4.28 |
|             | RMSE         | 9.25| 9.58| 9.93| 10.21| 10.47| 10.76| 10.03 |
| PR-LSTM     | MAE          | 3.91| 4.02| 4.09| 4.16| 4.22| 4.28| 4.11 |
|             | RMSE         | 9.12| 9.27| 9.51| 9.72| 9.89| 10.11| 9.60 |
| ConvLSTM    | MAE          | 3.98| 4.10| 4.19| 4.26| 4.34| 4.43| 4.22 |
|             | RMSE         | 9.18| 9.39| 9.62| 9.85| 10.13| 10.27| 9.74 |
| ST-ResNet   | MAE          | 3.89| 3.98| 4.06| 4.16| 4.25| 4.33| 4.11 |
|             | RMSE         | 9.08| 9.21| 9.45| 9.61| 9.83| 9.98| 9.53 |
| APR-LSTM    | MAE          | 3.75| 3.81| 3.86| 3.94| 4.06| 4.13| 3.93 |
|             | RMSE         | 8.78| 9.05| 9.14| 9.35| 9.50| 9.65| 9.24 |

(1) ARIMA [9]: The traditional model for time-series prediction, which cannot consider the spatial correlation.
(2) XGBoost [11]: The model has demonstrated outstanding performance in machine learning prediction areas. The static spatial correlations analysed from the correlation coefficients are reshaped as features and fed into the model.
(3) LSTM [14]: LSTM is widely used in time-series prediction. The input sequences of airport delay are fed into the LSTM-Seq2Seq framework with temporal attention mechanism.
(4) PR-LSTM [29]: The model including PageRank and LSTM-Seq2Seq framework, without temporal attention mechanism, which can analyse the effect of the temporal attention mechanism.
(5) Convolutional LSTM (ConvLSTM) [13]: The model combines CNN and LSTM at the bottom of the model, which is specifically designed for spatiotemporal sequences. The convolutional layers can extract spatial correlation and LSTM can capture the temporal correlations.
(6) ST-ResNet [23]: The model proposed by the team of Yu Zhen, employs the convolution-based residual neural network and summarises temporal properties such as closeness, period, and trend properties, to get the prediction results.

4.3 Experimental results

4.3.1 Model accuracy

Table 2 demonstrates the MAE and RMSE of different models at different prediction steps. The phenomena can be observed from the experiment results. Compared with other comparison models, the APR-LSTM model has the best accuracy and stability at each prediction step. It’s proved indirectly that the model can capture the spatial-temporal correlation greatly.
The metrics of all the models under the first prediction step are mainly similar because the delay status is relatively stable at the adjacent step.

The accuracy of deep learning models is better than XGBoost and ARIMA. The accuracy of ARIMA is worst in different step due to its poor ability of obtaining the spatial correlation and predicting the non-stationary time-series data. With the increase of the prediction step, the accuracy of all prediction models continues to decline. The error growth rate of XGBoost is significantly higher than in other deep learning models. Because the deep learning model can be combined with Seq2Seq, which can avoid the error accumulation phenomenon.

Compared with LSTM, the metrics of other deep learning models grow more slowly as the step size increases. This is because other deep learning models can mine dynamic spatial correlation and provides effective auxiliary information for the prediction model. Besides, the performance of PR-LSTM is better than ConvLSTM, because the structure of air traffic delay networks is non-Euclidean topology graph and unsuitable for using convolutional layer. ST-ResNet predicts the outputs dynamically by aggregating the three residual neural networks and introducing the closeness, period, and trend properties. Those temporal properties make the performance better than ConvLSTM.

The attention mechanism can significantly improve the prediction performances in each step. As the step size increases, the error trend is basically the same as the model without the attention mechanism, but the error rises more slowly.

### 4.3.2 Prediction curves

Figure 4 shows the prediction results of the different model on the different prediction step. The delay of Beijing capital airport from October 6th to October 9th are selected to observe the prediction metric of the delay prediction value under different models during holidays and daily periods. The prediction steps are 1 and 6 h.
After observing the prediction curves, the conclusions are shown as follows. All of the prediction models have high accuracy in single-step prediction, which can basically keep up with true delay values. The predicted values of XGboost have obvious lag and the prediction data is the smoothest. As the number of prediction steps increases, the prediction fluctuations of each model continue to increase. When the delay values are small, the tracking abilities of each model are better than the large delay. The APR-LSTM has the best tracking ability under the different prediction steps and can capture the extreme value of delay. Besides, the APR-LSTM model has the advancement in single-step prediction during the holidays, which can be informed of the serious delays in advance. Although the model proposed is the most sensitive to delay fluctuations, the ability for the wide range of delay fluctuations still needs to be improved.

4.4 Model interpretation

4.4.1 Spatial correlation analysis

PageRank can capture the airports whose delay has huge impact on entire network and analyse the important components of network. The sum of each airport’s PageRank value in 2016, denoted as sum$^\text{PR}(v_i)$, can be analysed the spatial correlation and the annual contribution of each airport on the network. The airports with large sum$^\text{PR}(v_i)$ are hub airports, which affect the delay status of other airports greatly.

The results show that only a few airports are important sections that have great impact on the delay of entire network. The importance of top 10 airport is shown in Table 3.

| No | Airport                  | ICAO  | sum$^\text{PR}(v_i)$ |
|----|--------------------------|-------|----------------------|
| 1  | Beijing Capital Airport  | ZBAA  | 528.92               |
| 2  | Shenzhen Bao’an Airport  | ZGSZ  | 463.12               |
| 3  | Shanghai Pudong Airport | ZSPD  | 375.55               |
| 4  | Guangzhou Baiyun Airport| ZGGG  | 361.92               |
| 5  | Hangzhou Airport         | ZSHC  | 356.64               |
| 6  | Chengdu Shuangliu Airport| ZUUU  | 312.62               |
| 7  | Kunming Changshui Airport| ZPPP  | 276.95               |
| 8  | Xi’an Xianyang Airport   | ZLXY  | 252.81               |
| 9  | Chongqing Jiangbei Airport| ZUCK  | 246.95               |
| 10 | Zhengzhou Xinzhieng Airport| ZHCC | 210.64               |

Besides, the PageRank vector PR($V_i$), can indicate the delay status of the network. The K-means cluster algorithm is used to identify characteristic delay states of the network and get the main network delay status. The vector used for cluster is PR($V_i$). Firstly, draw the figures by elbow method and silhouette method to determine the number of clusters, six is the best number of clusters. The cluster result and cluster centre are shown and visualised in Figure 5.

The cluster results indicate that the hub airports affect delay status of the entire network. In the figure, airports with bigger circles indicate the larger importance in the network.

4.4.2 Temporal attention mechanism analysis

We did the compared experiments to indicate the effect of attention mechanism. The results in Table 2 indicate improving the attention mechanism can significantly improve the prediction effect of each prediction step. With the step size increases, the error trend is basically the same as the model without attention mechanism, but the error rises more slowly. This is because the attention mechanism enhances the importance of long-term information and helps the decoding part learn more long-sequence information.

The prediction error grows slower with the increase of step size than the models without attention mechanism. The attention heatmap of input and output sequences is showed on Figure 6.

The figure reflects that the closer to the time of $t$, the weight of the prediction moment is greater. Besides, the weights of the last steps in the input sequence are relatively large. It means that there is a stronger correlation if the time is close to the decoder. When the distance between the encoded input sequence and decoded output sequence become too large, the correlation between them will be minuscule.

4.4.3 Flight characteristics and auxiliary features analysis

Flight delay can be affected by multi factors such as weather, flight characteristics, pre-flight delay propagation and the delay status of departure airport. The current research cannot take the dynamic delay status of airports into the flight delay prediction model, limiting the improvement of prediction accuracy. In this paper, the proposed model can obtain airport delay in advance, which can be taken into the flight delay prediction model to improve the prediction model accuracy. The stack Autoencoder and XGBoost model is proposed to predict the flight delay and has been verified the superiority of the model.

To investigate the impact of weather and dynamic spatial-temporal airport delay on the flight delay prediction, four datasets are constituted as F1, F2, F3 and F4. F1 only contains flight characteristics. F2 contains flight characteristics and weather data. F3 contains the flight data and dynamic spatial-temporal airport prediction delay and F4 contains three types
FIGURE 5  The distribution of cluster centre, including six network status. (1) The network in low delay: the proportion is 54.91%, (2) Beijing Capital Airport in high delay (12.03%), (3) Hangzhou Airport, Guangzhou Baiyun Airport, Shenzhen Bao'an Airport, Beijing Capital Airport and Shanghai Pudong Airport in medium delay (10.42%), (4) Shanghai Pudong Airport in high delay, Beijing Capital Airport and Shenzhen Bao'an Airport in medium delay (8.20%), (5) Shenzhen Bao’an Airport in high delay and Hangzhou Airport in medium delay (7.82%), (6) The network in high delay (6.62%).

of data. The prediction performance on different datasets are shown in the Table 4.

The prediction model on F1 dataset has the worst accuracy, because flight delay is affected by multi additional features. In addition, the MAE and RMSE of the model begin decreasing by introducing other features. The prediction error on F3 dataset is significantly lower than that on F2 dataset. It indicate that flight delay is strongly related to departure and arrival airport delay and the delay state of the whole network.

XGBoost is used to analyse the importance of features. Fscore indicates the number of times a feature is used to split dataset cross all trees. The larger Fscore value indicates that this feature is much more important than other features with lower Fscore.

It is obvious to see that the airport delay prediction result is the most important feature in the flight delay prediction model. Taking the spatial-temporal network delay and the airport prediction delay into consideration can effectively improve the accuracy of flight delay prediction. Therefore, the airport delay prediction can not only provide much more alert information for controllers and airlines, but also has great significant on improving the accuracy of flight delay prediction. It is essential for us to devote to the research of multi-step airport delay prediction. Besides, the analysis work indicates laterally that flight delay has strong spatial-temporal dependencies and is strongly affected by the delay status of airport and network, providing more research directions and basis for future research on flight delay prediction.
FIGURE 6  Temporal attention heatmap: The abscissa represents the step of input sequence and the ordinate represents the step of output sequence.

TABLE 4  The prediction performance on different datasets

| Dataset | MAE (min) | RMSE (min) |
|---------|----------|------------|
| F1      | 11.427   | 20.828     |
| F2      | 9.033    | 18.964     |
| F3      | 6.066    | 11.79      |
| F4      | 4.235    | 9.324      |

5 | CONCLUSION

Airport delay is affected strongly by spatial-temporal dependencies and other environmental features, resulting in more challenges in multi-step delay prediction. To tackle these challenges, with the spatial-temporal correlation and other auxiliary features considered, a sophisticated deep learning prediction framework named Attention PageRank LSTM-Seq2Seq (APR-LSTM) is proposed in this article. Firstly, the air traffic delay networks are modelled as weighted and directed graphs, and PageRank algorithm is employed to capture the dynamic spatial dependencies. Based on the spatial features of PageRank, the LSTM-Seq2Seq model is utilised to achieve multi-step prediction and the joint mining of the spatial-temporal dependencies via accepting the weighted input sequences of airport delays. Temporal attention mechanism is integrated into model to explore the relevance of different time steps. To validate the effectiveness of the proposed model, the results of comparative experiments indicate that the proposed model has better performances than other benchmark models among different intervals. We further analyse the spatial correlations of different airports, delay characterisations of networks, and temporal relevance of input sequences in detail. In addition, the prediction results are taken as features into the flight delay prediction model. Results demonstrate that the prediction airport delays are the most important features for flight delay prediction.

For the future work, we will focus on three parts. The first part is to improve the performances of prediction model, including accuracy, generalisation and robustness. The second part is to design a better prediction model to mining the dynamic spatial-temporal dependencies of air traffic delay network simultaneously. The last part is to take other important auxiliary features such as METAR weather and air traffic control information into the model to improve the performance of prediction model.

ACKNOWLEDGMENTS

The authors thank the associate editor and anonymous reviewers for their careful reading of the paper and very constructive comments that led to a substantially improved paper. This work was supported partially by the Key Research and Development Program of Guangdong (2020B0101050001), and partially by the Independent Project of State Key Laboratory of Industrial Control Technology (ICT2021A25).

ORCID

Chunyue Song https://orcid.org/0000-0003-2196-520X
Jie Zhang https://orcid.org/0000-0003-0564-0341

REFERENCES

1. Civil Aviation Administration of China. 2018 Statistical Bulletin of Civil Aviation Industry Development. May. 2019. [Online]. Available: http://www.caac.gov.cn
2. Krsti Simić, T., Babi, O.: Airport traffic complexity and environment efficiency metrics for evaluation of ATM measures. J. Air Transp. Manag. 42, 260–271 (2015)
3. Tay, G., Kalsi, H., Hornung, M.: A cluster-based approach for the assessment of air transportation networks in selected global regions. In: 17th AIAA Aviation Technology, Integration, and Operations Conf., AIAA, Menlo Park, CA (2017)
4. Sternberg, A. et al.: A review on flight delay prediction. arXiv:1703.06118 (2017)
5. Gopalakrishnan, K., Balakrishnan, H.: A comparative analysis of models for predicting delays in air traffic networks. In: Proceedings of the USA/Europe Air Traffic Management R&D Seminar, pp. 1–10. MIT Press, Cambridge, MA (2017)
6. Chen, D., et al.: Short-term traffic flow prediction of airspace sectors based on bayesian estimation theory. Xinan Jiaotong Daxue Xuebao/Journal Southwest Jiaotong Univ. 51(4), 807–814 (2016)
7. Pyrgiotis, N., Malone, K.M., Odoni, A.: Modelling delay propagation within an airport network. Transp. Res. Part C Emerg. Technol. 27, 60–75 (2013)
8. Wu Xiping, H.S., Hongyu, Y.: Analysis of air traffic characteristics and delay propagation based on complex network. J. Aeronaut. 38(S1), 113–119 (2017)
9. Rebollo, J.J., Balakrishnan, H.: Characterization and prediction of air traffic delays. Transp. Res. Part C Emerg. Technol. 44, 231–241 (2014)
10. HE Yang, Z.Q.:Jinfu, Z.: Airport flight delay prediction based on SVM regression. J. Civ. Aviat. Univ. China 036(001), 30–36.41 (2018)
11. Zhan, X., et al.: Multi-step-ahead traffic speed forecasting using multi-output gradient boosting regression tree. J. Intell. Syst. Technol. Planning, Oper. 24(2), 125–141 (2020)
12. Zhu, Q., et al.: Learning temporal and spatial correlations jointly: A unified framework for wind speed prediction. IEEE Trans. Sustain. Energy 11(1), 509–523 (2020)
13. Geng, X., et al.: Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. Proc. AAAI Conf. Artif. Intell. 33, 3656–3663 (2019)
14. Cui, Z. et al.: Deep stacked bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. arXiv:1801.02143, (2018)
15. Zhang, Z., et al.: Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. Transp. Res. Part C Emerg. Technol. 105, 297–322 (2019)
16. Sutskever, I., et al.: Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 4, 3104–3112 (2014)
17. Liao, B. et al.: Deep sequence learning with auxiliary information for traffic prediction. In: Proc. of the 24th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pp. 537–546. ACM Press, New York (2018)
18. Ma, X. et al.: Correlation analysis on delay propagation in aviation network. Command Inf. Syst. Technol. 9(4), 23–28 (2018), https://doi.org/10.15908/j.cnki.cist.2018.04.004
19. Gopalakrishnan, K., Balakrishnan, H., Jordan, R.: Clusters and communities in air traffic delay networks. In: Proc. of the 2016 American Control Conf., pp. 3782–3788. IEEE, Piscataway, NJ (2016)
20. Gui, G., et al.: Flight delay prediction based on aviation big data and machine learning. IEEE Trans. Veh. Technol. 69(1), 140150 (2020)
21. Wang, C., Wang, X.: Airport congestion delays and airline networks. Transp. Res. Part E Logist. Transp. Rev. 122, 328–349 (2019)
22. Du, W.B., et al.: Analysis of the Chinese airline network as multilayer networks. Transp. Res. Part E Logist. Transp. Rev. 89, 108–116 (2016)
23. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proc. of the 31st AAAI Conf. on Artificial Intelligence, pp. 1655–1661. AAAI, Menlo Park, CA (2017)
24. Cui, Z., et al.: Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. IEEE Trans. Intell. Transp. Syst. 21(11), 4883–4894 (2020)
25. Yu, R., et al.: Flight delay prediction for commercial air transport: A deep learning approach. Transp. Res. Part E Logist. Transp. Rev. 125, 203–221 (2019)
26. Carl, A.N.L., Meyer, D.: Deeper inside PageRank. Internet Math. 1(3), 335–380 (2004)
27. B, Z., Kun, Z., Peipei, L.: Evaluation method for node importance in directed-weighted complex networks based on PageRank. J. Nanjing Univ. Aeronaut. Astronautics 45(3), 429–434 (2013)
28. Murca, M.C.R., Hansman, R.J.: Identification, characterization, and prediction of traffic flow patterns in multi-airport systems. IEEE Trans. Intell. Transp. Syst. 20(5), 1683–1696 (2019)
29. Yuan, X., et al.: Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. IEEE Trans. Ind. Electron. 68(5), 4404–4414 (2021)
30. Yao, H., et al.: Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. Proc. AAAI Conf. Artif. Intell. 33(1), 5668–5675 (2019)
31. Zhu, L., Li, T., Du, S.: TA-STAN: A deep spatial-temporal attention learning framework for regional traffic accident risk prediction. In: Proc. Int. Jt. Conf. Neural Networks, pp. 1–8. IEEE, Piscataway, NJ (2019)