Airport Delay Prediction Based on Spatiotemporal Analysis and Bi-LSTM Sequence Learning

Hao Zhang
College of Control Science and Engineering
Zhejiang University
Hangzhou 310027, China
zh_zju@foxmail.com

Chunyue Song*
College of Control Science and Engineering
Zhejiang University
Hangzhou 310027, China
csong@zju.edu.cn

Hui Wang
College of Control Science and Engineering
Zhejiang University
Hangzhou 310027, China
hwangzju@zju.edu.cn

Chencheng Xu
College of Civil Aviation
Nanjing University of Aeronautics And Astronautics
Nanjing 211106, China
xuchch95@nuaa.edu.cn

Jinlong Guo
College of Control Science and Engineering
Zhejiang University
Hangzhou 310027, China
guojinlong@zju.edu.cn

***Abstract*—The air traffic network is the small-world system whose delay propagates among the entire network quickly. These frequent delays bring huge economic and resources loss. Therefore, accurately predicting the airport delay is essential for the airlines or air traffic controllers to adjust flight schedules. This paper proposes a multi-step deep sequence learning model (Bi-LSTM+Seq2Seq) to predict airport delay which considers the spatial-temporal correlation of other airports in the network. Firstly, the dataset is processed to analyze the temporal delay correlations of airports based on the complex network theory. The PageRank and K-means algorithm are used to cluster the behavior of the networks and to know the state of the entire network. Secondly, based on time series data about the current state of the network and delay relationship between airports, the Bi-LSTM+Seq2Seq model has been proposed and trained. Through the experiments, the proposed model has better accuracy and stability compared with other prediction algorithms.**

***Keywords*—delay network, spatiotemporal analysis, cluster, multistep prediction, Seq2seq+ Bi-LSTM**

I. Introduction

With the development of economy and technology, the air transportation industry develops rapidly and more and more passengers travel by air. However, the airport infrastructure and airspace are unable to meet the demand of the increasing number of flights, resulting in more and more serious flight delays. In the recent years, the rate of normal flights has continued to decline, standing at only 71.67% in 2017. Flight delays will cause a lot of economic losses and limit the development of air transport industry in China. Therefore, accurate prediction of the airport delay is significant for the air traffic control, airports and airlines, which can improve the operational efficiency of air transportation and provide passengers with a good flight experience.

The air transportation system is a complex system. The air traffic network consists of the airport nodes and air routes. Each airport does not exist independent and may be affected by other sections of the air traffic network. Many researchers at home and abroad have used complex network theory to model the air traffic network and analyze its structural characteristics. Considering the influence of other nodes in network can improve the accuracy of the prediction, especially for long-term traffic state prediction, T.Gilbert proposed a cluster-based approach for the assessment of air transportation networks. They find that the Chinese air traffic network have the small-world and scale-free characteristics and the connection between the airports in China is even closer and busier than in other countries [7], which means that the delay in a specific place will spread quickly and propagate among the entire network. K. Gopalakrishnan used the community detection algorithms such as Girvan and Newman or random walk to identify airports that have similar delays situation [1-2]. The shortest path length measurement, the degree centrality measurement, the clustering coefficient and sorting method based on feature vector can identify the important sections and analyze their spatial layout characteristics in the network. However, those methods are difficult to reveal the complex dynamics of the air traffic network and few studies take the influence of other airports into account.

In terms of air traffic prediction, the research objects can be divided into two categories from macroscopic to microscopic perspectives: one is predicting the delay of a single aircraft, and the other is predicting the average delay of the airport in a certain time period. In this paper, it is focused on the prediction of airports delay. The prediction methods are divided into two parts:

1) Prediction method based on statistical theory: the prediction models based on statistical theory focus on the spatiotemporal traffic information for predicting the perspective of probability. The common methods consist of Kalman filtering methods, Markov chains and Bayesian network. C.Weidong focused on the continuous airport situation of the same aircraft based on the Bayesian network, reflecting the distribution and spread of continuous airport delays[7]. L.Shanmei studied the probability distribution of airport traffic demanded from the perspective of uncertainty in order to obtain the congestion probability of the airport in the future[8]. N. Pyrgiotis explored the large-scale network and queuing model to explain the diffusion effect of the hub airport delay affects and the relationship between the relaxation time and delay of the associated flights ground crossing[9]. However, those researches don't take the spatial characteristic and the whole delay situation of network into consideration.

2) Prediction methods based on machine learning or deep learning: with the development of data mining technology, researchers begin to try to establish a data-driven machine learning model to predict airport delays. Many researchers use

This work was supported partially by the National Key Research and Development Program of China (No.2017YFA0700300), partially by the NSF under grant (61673342) of China, and partially by the Independent Project of State Key Laboratory of Industrial Control Technology (ITC1902).

the k-Nearest Neighbor, neural networks, SVM and random forests methods to predict the delays which focus on the characteristics in the time dimension and the average error equal to 10%-20%[13-18]. It's popular to model the complex characteristics of traffic data using the neural network. The deep learning model take the features such as weather conditions, holidays, social events or the number of flights into the model which decreases the error rate of model greatly[11,19-23].

Though many researchers focus on the airport delay prediction field and get great academic achievements, the proposed models mainly predict the single flight or airport delay problem, which ignores the spatial correlate relationship of other airports in the network. Besides, the recent time-series prediction models are mainly a single-step and short-term models which don't have guiding significance for air traffic control. In order to improve the shortcomings of the current air traffic network researches, this paper modifies PageRank algorithm to analyze the correlation of different airports and proposes a multi-step and long-term deep sequence prediction model to obtain the change of airport delay status, which takes the airport correlation and network status into the model.

The main framework of this paper is as follows: Chapter 2 introduces the information of dataset in this research. The dataset is processed to describe the air traffic delay networks. Chapter 3 uses the complex theory to explore the behavior of the air traffic delay network in order to find the spatial and temporal delay relationship of different airports. The correlation coefficient and PageRank algorithm can get the temporal delay relationship and K-means algorithm is used to cluster the delay behavior of the networks. In Chapter 4, the multi-step deep sequence learning model(BiLSTM Seq2Seq) is proposed and use the data of temporal attributes, the current state of the network and delay relationship between airports. The model proposed has better performance in long-term prediction, verifying the better accuracy than other methods.

II. Data description and Process

A. Description of data

The dataset used in this paper is collected by the VariFlight. The dataset contains the flight information from Jan. 1, 2015 to Dec. 31,2015, which consists of 3,523,471 pieces of actual flight operation data. Each flight operation data includes flight number, departure airport, arrived airport, actual/planned arrival time, actual/planned departure time, flight time and flight status. In this paper, only the data of domestic information is used, and the dataset is processed by deleting the operation data of the foreign flights, the canceled flights and flights whose delay time over 6 hours. Finally, 3,005,384 pieces of actual flight operation data are obtained.

B. Air network based on delay variable

Flight delay is calculated by the actual time minus the schedule time. If the delay is more than 15 minutes, this flight is considered as a delayed flight. This paper only considers the status of airports and routes, so the flight operation data are employed to calculate the average flight delay time of this route or this airport.

The air network based on delay variable can be considered as a graph $G = (V, E)$, where V is the set of nodes and E is the set of the edges between them. $V = \{v_1, v_2, ..., v_n\}$ is represented as airports and $n = |V|$ is the number of network's nodes. $E = \{e_1, e_2, ..., e_n\}$ is represented as the route and $m = |E|$ is the number of network's edges. As for airport v_i and v_j , each edge is represented as an origin-destination (OD) pair (v_i , v_j) if airport v_i and v_j have flight. Let the weight on edge (v_i , v_j) be denoted by $\omega_{i,j}$. We can build the network as the adjacency matrix A:

$$A_{i,j} = \begin{cases} \omega_{i,j} & (v_i, v_j) \text{ has delay} \\ 0 & \text{else} \end{cases} \tag{1}$$

When $i \neq j$, $\omega_{i,j}$ is defined as the average flight delay time of this route and when $i = j$, $\omega_{i,j}$ is defined as the average flight delay time of this airport. When the traffic delay network is constructed, the origin-destination (OD) pairs which have at least 10 flights a day on average is taken into consideration. The processed network data contains a total of 114 airport nodes and 2360 routes. The airport delay is divided into the inbound and outbound average delay time d_{in} and d_{out}. Every hour an air traffic network are generated in the past 365 days, bringing a total of 365*24=8760 networks.

III. Spatiotemporal Network Analyse

The research in the past considered the temporal and spatial correlation of airport rarely. In fact, there is a strong spatial-temporal correlation between airport in air network based on delay variable. This part mainly uses the complex network theory to analyze the air network to get the relationship between different airports in spatiotemporal feature and fully know the structure of the delay networks.

A. Airport spatiotemporal correlation analysis

In this part, the correlation of the delay time series between the airports is analyzed, which can describe the correlation and the linear relationship of airport delay between two airports' delay at different times. the cosine correlation and the Pearson correlation coefficient are mainly used to analyze the time series correlation between 114 airports.

1) The temporal autocorrelation coefficient ρ_m^i : The autocorrelation coefficient ρ_m^i indicates the temporal correlation of between the observation at time t and the hysteresis m periods about the airport i. The equation of autocorrelation is as follow:

$$\rho_m^i = \frac{\sum_{i=1}^{n-m}\left(D_t^i - \overline{D^i}\right)\left(D_{t-m}^i - \overline{D^i}\right)}{\sum_{i=1}^{n}\left(D_t^i - \overline{D^i}\right)^2} \tag{2}$$

D_t^i is the delay of airport i at the time t. D^i is the time series vector as $\{D_{t_1}^i, D_{t_2}^i, ..., D_{t_n}^i\}$. n is the length of time-series D^i. The value of is range from -1 to 1.

2) The temporal correlation coefficient $\cos\theta_{i,j}$

$cos\theta_{i,j}$ is the coefficient that characterizes the spatial similarity of airport i and j. It's defined as:

$$\cos\theta_{i,j} = \frac{\sum_{t=1}^{m} D_t^i \cdot D_t^j}{\sqrt{\sum_{t=1}^{m}\left(D_t^i\right)^2}\sqrt{\sum_{t=1}^{m}\left(D_t^j\right)^2}} \tag{3}$$

The value of is range from -1 to 1. If $cos\theta_{i,j} = 1$, it means the time series of airport i and$j$ are completely similar about and If $cos\theta_{i,j}$ = -1 ,it means the time series are completely different. If the delay coefficients are more than 0.7, it means that this OD pair has strong delay correlation.

After the analysis, we can find temporal correlation discipline that $m = 6$, which means that the six piece of time has a strong connection with the airport delay. Besides, the most of the spatial coefficients are distributed between 0 and 0.2, which indicates that most of the airport in China have no significant spatial correlation. It is because there is no directly connected flights between many airports. It can be found that these airports which have strong delay correlation are geographically closed. Because their operating environment and weather conditions are very similar, which shares the same sectors. Therefore, the delay between airports with the same operating environment, geographical location and weather conditions is more relevant. the spatiotemporal characteristics in the prediction model should be considered. The airport delay coefficient reflects the relationship between individual airport delays in time temporal features.

B. *PageRank*

PageRank is a iterative algorithm used by Google to identify the level/importance of a web page. PageRank is the algorithm for judging which airport in the delay state have a huge impact on the entire air network. It also can analyze important components of the network structure. The importance of the web page is analyzed by the number and weight of hyperlinks between web pages. It’s based on the theory of “When a page is connected to more pages, it is an important node. When a page is connected by a high-quality page, the more important the page is” to judge the importance of all pages. In PageRank, the importance of pages is measured by the probability of $PR(v_j)$ (usually called the rank value) accessed by the web page. Assuming that the viewer is browsing a web page, he will click on the hyperlink with a probability d (the damping coefficient) or click new page with probability 1- d. Under this model, the probability $PR(v_j)$ is defined as:

$$PR\left(v_j\right) = \frac{1-d}{n} + d\sum_{i=1}^{m} \frac{PR\left(v_i\right)}{C_{out}\left(v_i\right)} \tag{4}$$

n is the total number of the pages. $C_{out}(v_i)$ is the out-link number of the page v. m is the in-link number of the page u.Give each page an equal initial value, then use equation $PR(v_j)$ to iterate until the calculated value converges, The principle is similar to the Markov chain. However, traditional PageRank algorithm only can be used to calculate the unweighted network.

This algorithm is strengthened to calculate the directional weighted network. $S_{out}(v_i)$ is the out-strength of node v_i. $V_{out}(v_i)$ is the collection of nodes that has edge to node v_i. The new definition is as below[7]:

$$S_{out}\left(v_i\right) = \sum_{v_j \in V_{out}\left(v_i\right)} \omega\left(v_i, v_j\right) \tag{5}$$

$$PR\left(v_j\right) = \frac{1-d}{n} + d\sum_{i=1}^{m} \frac{\omega\left(v_i, v_j\right)}{S_{out}\left(v_i\right)} PR\left(v_i\right) \tag{6}$$

PageRank is a Markov transfer process in essential. Results are obtained by iterating the state transition matrix continuously until error less than 10^{-6}. The state transition matrix is $M_{i,j} = \dfrac{\omega\left(v_i, v_j\right)}{S_{out}\left(v_i\right)}$ if (v_i, v_j) have link.

The improved PageRank can ameliorate the irrationality of the average allocation of traditional algorithms as well as the value of important nodes. In this paper, there are 8760 of the PageRank vector $\overrightarrow{PR}$ obtained via using the improved algorithm and get the most important airport in the delay network. The importance of top 12 airport are shown as follow:

TABLE I. THE MOST IMPORTANT AIRPORT

	City	Airport		City	Airport
1	Beijing	ZBAA	2	Shenzhen	ZGSZ
3	SH Pudong	ZSPD	4	Hangzhou	ZSHC
5	Chengdu	ZUUU	6	Guangzhou	ZGGG
7	Chongqing	ZUCK	8	Kunming	ZPPP
9	Nanjing	ZSNJ	10	Wulumuqi	ZWWW
11	Xian	ZLXY	12	SH Hongqiao	ZSSS

The distribution of the important airports in China is as follow: we can see the important airport almost locate in the east of China and the capital of province. ZBAA\ ZGSZ\ZSPD are the most important airports in the network and once happening the serious delay, the whole network will become terrible. It’s easy to get that the importance is related by the level of economic development, because the huge airports have a lot of flights flying to the other airports.

Fig. 1. The distribution of the important airport in the network

C. *Network cluster*

The PageRank vector features discussed in the previous section can be used to compare networks and to cluster similar networks. It means that two observed air traffic delay networks can be judged the delay state similarity if these features are close to another. The Euclidean distance between the feature vectors can be used to measure distance. The feature studied vectors included: PageRank score weighted by the total delay in the system: $State_t = \left(\sum_{i=1}^{n} D_t^i\right)\overrightarrow{PR}$

The general cluster algorithm concludes k-means or k-medoids algorithms, which are used to determine clusters of similar graphs[2]. K-means algorithm is an iterative clustering algorithm. The first step is randomly selecting K points as the initial cluster center, then calculate the distance between each point and each cluster center, and assign each object to the most nearest cluster center. The Euclidean distance between the feature vectors was utilized as the objective function. The cluster centers and the objects assigned to them represent a cluster. The cluster center will be recalculated based on the existing objects in the cluster. This process will be repeated until a termination condition is satisfied such as cluster centers no longer change or squared error is minimum.

The number of clusters is evaluated using various criteria, including the sum of distances to the nearest cluster centroids .

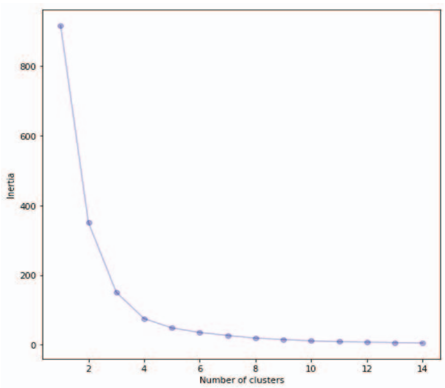

Fig. 2. The iteration probability of the airport delay coefficient

In this paper, those 8760 graphs are analyzed via the k-means. The summation of distances to the nearest cluster centroids is used to decide the number of the cluster. Although it is ideally to get the smallest summation, six is the best number of clusters as shown in the picture because when the data is divided into many clusters, time and computational complexity would rise exponentially. So It's obvious that six is the best number of clusters in the picture. the cluster number is set as 6to get the cluster center and the elements in clusters. The result and cluster centers are as follow:

TABLE II. THE CLUSTER CENTER OF DELAY NETWORK

	center	The elements in cluster
1	ZBAA High Delay	685
2	ZSHC/ZUUU/ZGSZ/ZBAA Medium Delay	913
3	ZSPD High Delay ZBAA/ ZGSZ Medium Delay	718
4	ZGSZ High Delay ZSHC Medium Delay	1054
5	Total network delay	580
6	Total network low delay	4810

IV. DELAY NETWORK PREDICT MODEL

A. Prediction model

The air traffic network is a nonlinear and complex system, so only using linear regression may be not useful. Delay prediction refers to statistical analysis of historical time series data, and speculates on future development trends. So prediction is critical that many parametric and non-parametric methods are proposed such as ARIMA, Random Forest. ARIMA is not suitable for large-scale traffic data. In general, short-term predictions based on time series are better than long-term predictions. However, the short-term prediction is out of significance, so more and more research are focusing on long-term predictions.

The LSTM(Long Short-Term Memory) neural network is an improved deep learning algorithm based on RNN, especially for processing time series data. LSTM algorithm contains forget gate, input gate, output gate and memory unit donated as f_t, i_t, o_t. The input gate accepts the input information and updates the state of the memory unit according to the input gate condition. The forget gate determines the discarded information according to the specific condition, and the output gate determines the output content according to the input information and the memory unit[6]. The calculation formula of each variable is as follows:

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \tag{7}$$

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \tag{8}$$

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o\right) \tag{9}$$

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \tag{10}$$

$$h_t = o_t \tanh\left(c_t\right) \tag{11}$$

where W_{xf}, W_{xi} and W_{xc} are the weight matrices mapping the hidden layer input to the three gates and the input cell state, while W_{ho} , W_{hi} , W_{hf}, W_{hc} are the weight matrices connecting the previous cell output state to the three gates and the input cell state. The b_f , b_i , b_o , b_h are four bias vectors. The σ is the gate activation function, which normally is the sigmoid function or hyperbolic tangent function. c_t is the cell output state and h_t is the layer output.

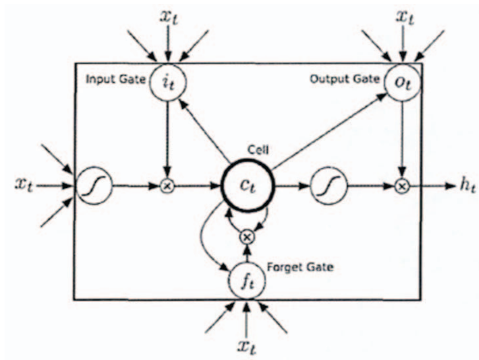

Fig. 3. The LSTM architecture

The traditional LSTM models are predicted from the front to the back, so the subsequent data points will be more important than the previous ones, which will often miss the information of many long associated data points. Bidirectional -LSTM(Bi-LSTM) improves this defect by performing two forward and backward LSTM training for a training sequence. The Bi-LSTM equation is defined as follow:

$$s_t = f\left(Ux_t + Ws_{t-1}\right) \tag{12}$$

$$s_t^{'} = f\left(U^{'}x_t + W^{'}s_{t-1}\right) \tag{13}$$

$$o_t = g\left(V s_t + V' s_t'\right) \quad (14)$$

s_t is the forward calculation of LSTM hidden layer status and s_t' is the backward calculation of LSTM hidden layer status. The f and g is the gate activation function. In this paper, the Elus function is used and defined as follow:

$$f(x) = \begin{cases} \alpha\left(e^x - 1\right) & x < 0 \\ x & x \geq 0 \end{cases} \quad (15)$$

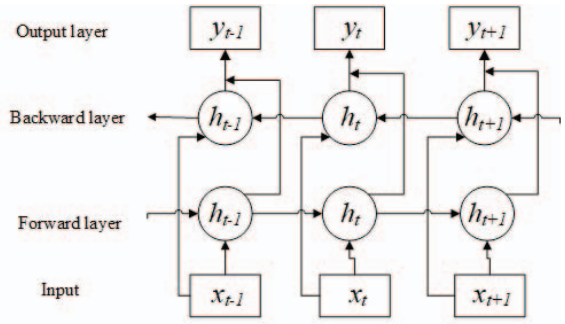

Fig. 4. The Bi-LSTM architecture

The sequence to sequence model is the most popular algorithm and focus on "many-to-many" prediction problem. It was originally developed for machine translation problem, although it has been proven successful at related sequence-to-sequence prediction problems such as text summarization and question answering. This approach involves two part: one to encode the source sequence called encoder and a second to decode the encoded source sequence into the target sequence, called the decoder[10,20,23]. This algorithm has some advantages: 1) the encoder and decoder can accept input data with varying length and the model is more flexible. 2)This model can learn the temporal relationship better during the training process. So it has the strong robustness.

The input sequence of this model is [D_{t-p}, D_{t-p+1},…, D_t] and [$PR(t-p)$, $PR(t-p+1)$,…, $PR(t)$] which can reflect the spatial relationship in air traffic network. The output sequence is [D_t, D_{t+1},…, D_{t+q}]. The vector D_t is $[D_t^1, D_t^2, ..., D_t^n, State_t]$, which D_t^i is the delay of the airport i and the $state_t$ is the network delay state at the time t. p is the number of the input step and the q is the predict step. Otherwise, the airport delay coefficient matrix are considered in this model.

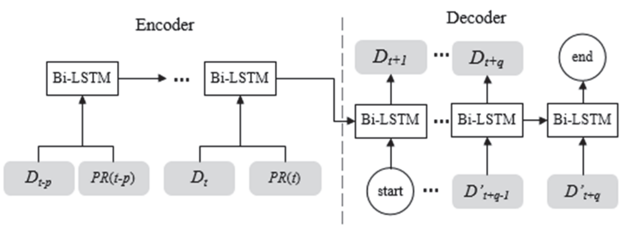

Fig. 5. The architecture of seq2seq prediction mode

In order to evaluate the performance of the model, this paper use the mean absolute percentage error (*MAPE*) to verify, which is defined as:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{D_t - \widehat{D}_t}{D_t}\right| \times 100\% \quad (16)$$

where D_t is the actual airport delay and D_t is the predicted airport delay at time t. N is the number of predicted value. *MAPE* is a dimensionless metric which reflects the error level and confidence.

B. Prediction Experiment

The experimental system is win10 64-bit and the python is 3.7.0 and the keras version is 2.0.8. In order to evaluate the performance of the model, this experiment uses the data of the air delay network paper and the behavior of delay network. The data interval is 1h and the data set contains 8760 pieces network. China's aviation has two sets of schedules and the most serious delay occurs in Jun, July, August and September. On the other hand, delay continued to increase at 6 am and would peak between 19:00 and 22:00.So data in the time series belonging to holidays and weekends is marked as the additional feature and released for training.

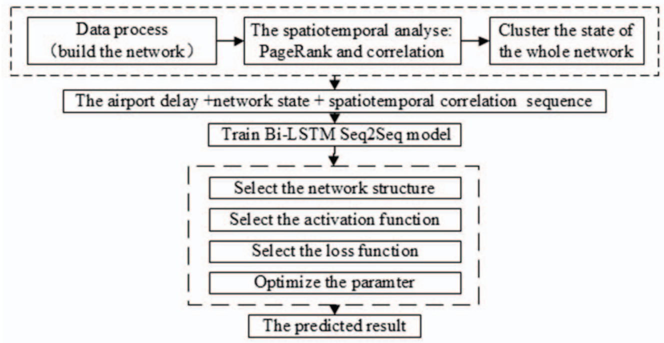

Fig. 6. The process of predict model

The length of input sequence is $p = 6$ because in the part 2, we find that temporal correlation discipline is $m = 6$, which means that the six piece of time has a strong connection with the airport delay. The length of output sequence is $q = 3$ because the loss will increase as the length of output increase. 80% data is used as training set and the other is used for testing. A mini-batch stochastic gradient descent (SGD) method is used to update the training parameters. The mini-batch size is 128 and the hidden state size is 64.

In this experiment, traditional LSTM and the model is compared without the information of airport delay coefficient matrix. We treated the prior prediction as the next step's input. The results are as follow:

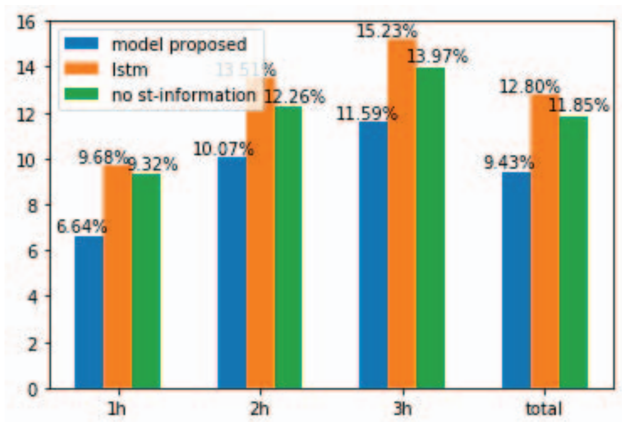

Fig. 7. The comparison of different model

C. Result analysis

The experiment shows that the Bi-LSTM+Seq2Seq has the lowest MAPE error. Compared with the traditional LSTM method, the Bi-LSTM has better performance with the total

MAPE 9.43%. As the number of prediction steps increases, the prediction error increases significantly lower than LSTM. The reason is that the subsequent data points will be more important than the previous ones, which will often miss the information of many long associated data points in traditional LSTM model. Bi-LSTM does two forward and backward LSTM training for a training sequence to solve disadvantages. The MAPE of the model without spatiotemporal information is 11.48% and more than the model proposed. As the prediction time expands, the error grows the slowest compared by other models. The spatiotemporal information is able to improve the accuracy and prediction value has better tracking performance and the delay propagate through the network.

V. CONCLUSION

The air traffic network is a complex system whose delay will propagate among the entire network quickly. In this paper, air traffic delay networks are firstly built and analyzed by the processed dataset. Secondly, The temporal relationship are analyzed and the PageRank algorithm can select the important airport. Then, the K-means algorithm cluster the same behavior of the networks which can reflect the spatial relationship. Finally, the multi-step deep sequence learning model based on time series data about the current state of the network and spatiotemporal relationship between airports are proposed. The proposed model has better accuracy and stability than other algorithms. In the future, much more information such as weather or air traffic control information will be taken into the model to improve the performance and accuracy of the model.

REFERENCES

[1] K. Gopalakrishnan, H. Balakrishnan and R. Jordan, "Clusters and communities in air traffic delay networks," 2016 American Control Conference (ACC), 2016, pp. 3782-3788

[2] W. Xiping , Y. Hongyu , H. Songchen,"Analysis of properties and delay propagation of air traffic based on complex network," Acta Aeronautica et Astronautica Sinica, 2017, pp. 38(S1): 721473

[3] K. Gopalakrishnan and H. Balakrishnan, "A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks," Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017),2017

[4] Q. Shao, Y. Zhu, M. Jia, "Analysis of flight delay propagation based on complex network theory," Aeronautical Computing Technique, 2015, pp.45(4): 24-28

[5] Z. J. Hanley, "Delay Characterization and Prediction in Major U.S. Airline Networks," Master's thesis, Massachusetts Institute of Technology, 2015

[6] T. Gilbert, K.Henri, H.mirko "A Cluster-Based Approach for the Assessment of Air Transportation Networks in Selected Global Region", 17th AIAA Aviation Technology, Integration and Operations Conference, 2017,pp:10.2514/6.2017-3773

[7] Z.Kun, L.Peipei, Z.Baoping and H.Manyyu, "Evaluation Method for Node Importance in Directed-Weighted Complex Networks Based on PageRank", Journal of Nanjing University of Aeronautics & Astronautics, 2013.03.001.Vol.45.No.3 ,pp:429-434

[8] C.Weidong and H.Guoguang,"Bayesian networks analysis for sequence flight delay and propagation",Journal of Computer Applications,2009,pp.29(2):606-610

[9] L. Shanmei, Z. Yuandi, "Forecasting of Airport Network Traffic Congestion Propagation Based on Division of Airport Subareas", Journal of Transportation Systems Engineering and Information Technology, 2017, pp.17(4):207-213

[10] N. Pyrgiotis, K.M. Malone, A. Odoni,"Modelling delay propagation within an airport network," Transportation Research Part C: Emerging Technologies, 2013, pp. 27: 60-75.

[11] B. Liao, J. Zhang, C. Wu, "Deep sequence learning with auxiliary information for traffic prediction," Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining , 2018, pp. 537-546

[12] H. Haoyue, "Research on prediction of flight delay based on big data methods," Master's thesis, Nanjing University of Aeronautics and Astrinautic,2017

[13] A. Sternberg, J. Soares, D. Carvalho and E. Ogasawara, "A review on flight delay prediction," 2017

[14] S. Khanmohammadi, S. Tutun and Y. Kucuk, "A new multilevel input layer artificial neural network for predicting flight delays at JFK airport," Procedia Computer Science, 2016, pp.95:237-244

[15] Y.J. Kim, S. Choi , S. Briceno, " A deep learning approach to flight delay prediction," Digital Avionics Systems Conference. IEEE, 2016. pp.1-6

[16] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," Transportation Research Part C, 2014, pp. 231–241

[17] H. Cheng, Y.M. Li, Q. LUO, "Study on flight delay with C4.5 decision tree based prediction method," System Engineering-Theory & Practice, 2014, pp.34(s1): 239-247

[18] E. Demir, V.B.Demir, "Predicting flight delays with artificial neural networks: Case study of an airport," Signal Processing & Communications Applications Conference, 2017

[19] Bin Yu,Zhen Guo,Sobhan Asian,Huaizhu Wang,Gang Chen. "Flight delay prediction for commercial air transport: A deep learning approach", Transportation Research Part E,2019,125.

[20] H. Yang, Z. Jinfu and Z. Qinyan, "Airport flight delay prediction based on SVM regression," Journal of Civil Aviation University of China, 2018, pp.36:30-41

[21] W. Renbiao, L. Jiayi, Q. Jingyi, "Flight delay prediction based on dual-channel convolutional neural networks," Journal of Computer Applications , 2018, pp.38(07):2100-2106+2112

[22] T.Tao, Z.Xi, M.Bo and Z.Fan, "Abnormal time series data detection of gas station by Seq2Seq model based on bidirectional long short-term memory", Journal of Computer Applications, ISSN 1001-9081, 2019,39(3) : 924 -929

[23] Y.Bin and Y.Haiteng ,"Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting", Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18),2018, pp:3634-3640