2023 International Conference on Cyber-Physical Social Intelligence (ICCSI) | 979-8-3503-1249-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCSI58851.2023.10304027

Research on Flight Delay Propagation Prediction Method Based on Transformer

Jingyi Qu*
Tianjin Key Laboratory of Advanced Signal Processing
Civil Aviation University of China
Tianjin, China
* Corresponding author: qujingyicauc@163.com

Lin Zhang
Tianjin Key Laboratory of Advanced Signal Processing
Civil Aviation University of China
Tianjin, China
Beriaaa2022@163.com

Shixing Wu
Tianjin Key Laboratory of Advanced Signal Processing
Civil Aviation University of China
Tianjin, China
17864270365@163.com

Abstract*—In recent years, flight delay has become a problem that needs to be solved urgently. The problem of flight delay propagation prediction is studied in this thesis, which is the phenomenon of large area flight delay caused by an airport delay. The analysis and prediction of flight delay propagation in advance can assist civil aviation departments in controlling the flight delay rate and reducing the economic loss caused by flight delays. Firstly, the national flight data is cleaned and filtered to construct a dataset of flight chain affected by flight delay. At the same time, errors or missing values in the data are removed, and data related to flight delay are selected for subsequent analysis and modeling. Modeling analysis is performed on the issue of flight delay propagation, and a model of flight delay propagation is constructed to analyze the impact of flight delay propagation in airport network transmission. The model fully considers the interrelationship between flights, treating the flight chain as a whole rather than individual flights. This will be more conducive to exploring and understanding the causes and impacts of flight delay. Subsequently, in terms of constructing the dataset of flight chain structures, the method of data preprocessing and triplet flight chain dataset contains a series of flight chain data that can be used to predict the scope of flight delay propagation. Secondly, according to the features of flight delay, the Transformer model was adjusted and improved to enhance the accuracy and efficiency of prediction. The network architecture is restructured by injecting new convolutional pooling modules into the original network and adjusting the input and output dimensions of the network to make the overall model more tailored to flight delay prediction. Last but not least, the results of the experiments are showed that the modified model has low complexity, shorter time and good real-time performance. It makes prediction and decision faster, improves the real-time performance and better meets the needs of practical applications with an accuracy of 90.3%.

***Keywords*—Flight delay propagation prediction; Flight delay propagation model; Flight chain structure dataset; Transformer**

I. INTRODUCTION

In the flight delay problem, due to the influence of various factors, the delay of the previous flight may cause the delay of many downstream flights, and then cause widespread delays. The problem studied in this paper is the flight delay propagation prediction, that is, the chain effect prediction of flight delay. The purpose of this paper is to analyze the impact of flight delay propagation prediction on airport network propagation and forecast the degree of delay of each flight by inputting relevant data. We can then allocate reasonable relaxation time according to the predicted degree of delay can absorb the impact of upstream flight delay, alleviate flight delay to a certain extent, and provide theoretical support for relevant departments to optimize flight planning and control flight delay.

At present, domestic and foreign academicians have done some research on the problem of airport delay. Traditional algorithms such as decision tree and neural network are used in the existing methods. Decision tree is a classifier based on tree structure. Khanmohammadi et al. proposed a multilevel input-layer neural network algorithm based on C4.5 decision tree method for flight delay prediction [1]. However, because the decision tree algorithm is a greedy algorithm, it will overfit the data and produce overly complex models. Neural network carries out distributed and parallel information processing by imitating the synaptic connection of the brain [2-4]. The purpose of neural networks is to establish mapping relationships between inputs and outputs by learning complex relationships between data. Among them, RNN, Recurrent Neural Network [5] and Recurrent neural network, such as LSTM, Long-Short Term Memory [6], are a kind of neural networks that can process sequential data, among which LSTM can also deal with long-term dependence. Noborut et al. suggested to use shallow artificial neural network to predict flight delays [7]. Wu Renbiao et al. gather both the superiorities of CondenseNet and CBAM (Convolutional Block Attention Module) and use channels and spatial attention mechanisms to enhance the transmission of deep information from the network structure, thus effectively improving the network performance [8]. CBAM-CondenseNet network and SimAM-CNN-MLSTM (SimAM-CNN-MogrifierLSTM) network [9]. CBAM combines superiorities

979-8-3503-1249-2/23/$31.00 © 2023 IEEE

of CondenseNet and CBAM. Global maximum pooling layer and global mean pooling layer are used in CBAM in spatial dimension to compress and aggregate information into weight arrays, which enhances the information transmission of deep network structures. CondenseNet is a dense network based on convolutional neural networks. Excellent feature extraction ability and high computational efficiency are possessed by it. CNN convolutional layer was used in SimAM-cnn-mlstm to initially extract spatial characteristics of flight chain dataset. After each convolutional layer, SimAM attention module is added to synchronize channel and spatially weighting key feature information. Finally, MogrifierLSTM is used to enhance the ability of context modeling, better learning time sequence features. However, the neural network has some shortcomings, such as disappearing gradient, insufficient long-term memory ability, unable to deepen network structure and low precision.

In order to solve the above network deficiencies, Transformer network is used to realize the forecast of flight delay propagation [10]. In the task of flight delay propagation prediction, Transformer network can receive a set of time series data as input to predict the possibility of future delay. In this paper, encoders in Transformer networks are used to convert the sequence data into feature representations, combined with the Full-Connection layer and Softmax to output the predicted results. The improved Transformer model can handle long sequence data well, and parallel computing can save computing time and realize fast parallelism by using the Self-Attention mechanism. Moreover, Transformer can be increased to a very deep depth to fully explore the characteristics of the model and improve the accuracy of the model.

II. DATASET CONSTRUCTION

Flight data usually includes flight number, takeoff and landing time, departure place and destination, flight status and other information, while airport data includes initial flight delay status, secondary flight delay status, and the relationship, etc., which requires data construction in the process of preprocessing in this paper. The data was provided by the Administration of China Eastern Air Traffic Management. Since the original data of civil aviation operation is incomplete, inconsistent and contains noise, it is essential to pretreatment the raw data of airport delay. The process of data preprocessing mainly includes data cleaning, data fusion and data coding.

A. *Flight delay and data preprocessing*

In the steps of flight data cleaning, invalid data should be removed, and missing values should be filled. Then, the data of flight chain should be corresponding to time and place to better explore the relationship between them. In the data fusion stage, the impact of the previous flight is one of the important factors affecting the on-time performance of the flight. Therefore, the accuracy of flight delay prediction can be improved by integrating the current flight data with the preceding flight chain data. In this paper, the flight data link is taken as the input data, with each piece of data representing a time step, and the states of the first two pieces of data are influences on the states of the third piece of data. Therefore, the flight chain data can be used as input and the delay states of the three flight chains as output, and the Transformer model can be used to train this sequence to predict the delay of the third piece of data. The process of data encoding includes two main steps: feature extraction and feature encoding. Feature extraction is the process of converting the original flight chain data into computable feature vectors. Feature coding is the process of encoding extracted feature vectors. The flight data contains all discrete data and numerical data. CatBoost coding is carried out for discrete feature data, and Min-Max normalization coding is carried out for numerical feature data. Through data coding, flight data is converted into the form of input model, which provides the basis for the subsequent feature extraction and classification prediction.

B. *Construction of Flight Chain Dataset*

The airport where the same aircraft takes off for the first time within a certain time range is defined as the cascade 1 airport. The airport where the aircraft arrives from the cascade 1 departure airport for flight task 1 is called the cascade 2 airport, also known as the cascade 1 arrival airport or the cascade 2 departure airport, and so on to form the relationship of flight chain. However, when a flight in the current sequence is delayed, its own delay may lead to the inability to continue to fly subsequent flights, and may lead to the congestion of crew resources, which may affect its subsequent flights or even other flights, causing a wider spread of flight delays, which forms the impact of flight delays, and also makes the impact of flight delays have the characteristics of space-time distribution. This phenomenon can be illustrated by a single aircraft performing three flight missions in a row. The flight chain model is shown in the Fig. 1.

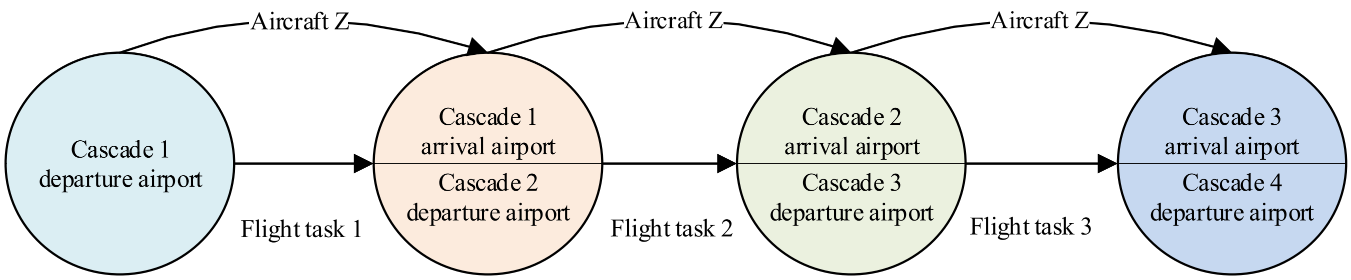

Fig. 1. Flight chain model.

The steps of flight data link construction are as follows:

1) Feature expression extraction: Feature attributes related to flight delay and contagion relationship are extracted from the original dataset. Based on such feature attributes, flight status can be expressed in a normative form in this paper;

2) Constructing data triples: For each flight chain data, this paper can represent it as a triple form. For example (f11, f12, f13), where f11 denotes the flight data of the primary flight, which contains all the characteristic attributes and the relations of the primary flight, and the flight information of the tertiary flight data is arranged in the form of a sequence to form a triad, which together constitute a chain of flight data, whose corresponding delay status information is calculated by labels and matched with the triad in the form of The corresponding delay status information is calculated by labels and matched with the data triad in the form of a triad, and the label triad is represented as (L1, L2, L3), and the construction process is shown in Fig. 2.

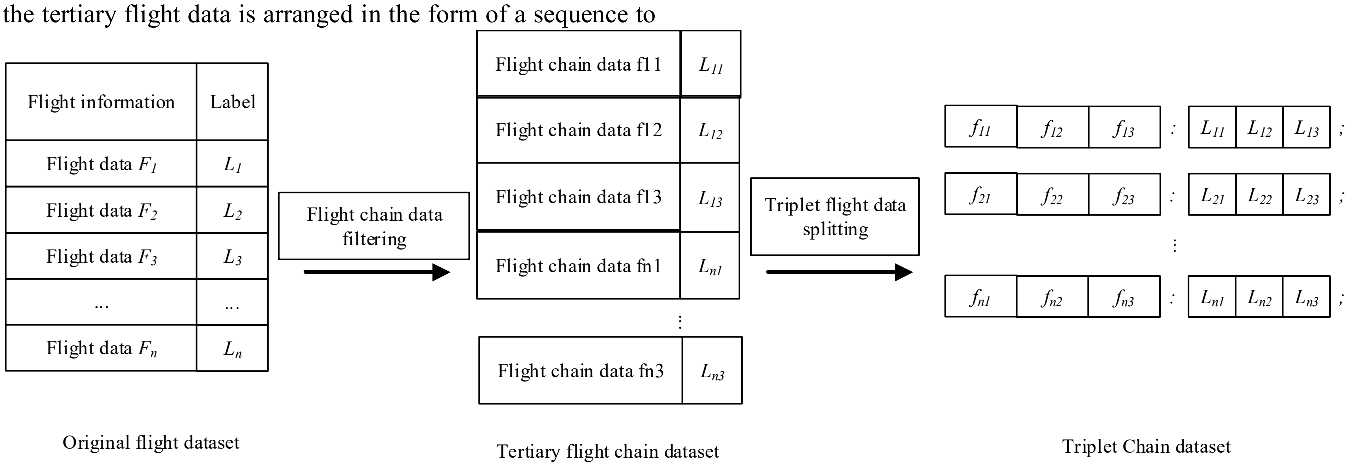

Fig. 2. Schematic diagram of triplet construction of flight data.

Flight link data is innovatively reconstructed by this paper in the form of triples to better express the characteristics of flight link data. This method can not only avoid the loss of some key features in the process of modeling and training, but also extract and process the time-related features more easily, so that the model can forecast the overall level of delays in flight chain more accurately.

TABLE I. FLIGHT DELAY CLASSIFICATION.

Delay Level	**Delay time T/min**	**Delay Level Classification**
No delay	$T \leq 15$	0
Minor delay	$15 < T \leq 60$	1
Moderate delay	$60 < T \leq 120$	2
High delay	$120 < T \leq 240$	3
Significant delay	$T > 240$	4

According to the definition of flight delays in the "Regulations on Normal Flight Management", we subdivide the flight delays into five classes and classify the delay classes according to the criteria of different classes, as shown in TABLE I. . In order to predict the flight delay class, a Softmax classifier is used to assign a delay class label to each flight by considering the variance between the scheduled arrival time and the real-time arrival time.

III. FLIGHT PROPAGATION MODEL

Transformer is a sequence-to-sequence model, that can model and predict flight delay wave and prediction tasks based on a network structure with better parallelism and in line with existing GPU frameworks [10]. Using Transformer model can better handle sequence information in historical data and capture the interaction between data, as well as long sequence data. In addition, the Transformer model can also perform a multi-headed self-attentive mechanism, which can focus on multiple positions of the input sequence at the same time to better capture the features and complexity of the time series. In flight delay prediction, historical data is input into the Transformer model in chronological order for training, and then the model is used for prediction. For the consideration of flight delay propagation prediction, the predicted delay results can be input into a subsequent model for analysis to determine the delay probability of other flights, and its parallel computing structure with multiple head mechanisms also makes it much more efficient in terms of computation.

The Transformer consists of several encoders and decoders. The core element in the encoder, Self-Attention, is to learn a weight for each feature of the input vector, measuring the relationship between each feature. The Self-Attention mechanism is computed as follows:

A. First, the Query, Key and Value vectors of each input element are obtained by three linear transformations. These three vectors are calculated as Equation (1)-Equation (3):

$$Q_i = W_Q x_i \tag{1}$$

$$K_i = W_K x_i \tag{2}$$

$$V_i = W_V x_i \tag{3}$$

Where W_Q, W_K, W_V are all matrices of d×d which are used to map each input element into the Query, Key and Value spaces, respectively.

B. Then, the weights are obtained by calculating the similarity of the Query vector to all Key vectors. The Scaled Dot-Product Attention mechanism is used here, and the formula is:

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d}})V \tag{4}$$

Among them, QK^T denotes the inner product between the Query vector and the Key vector, $\sqrt{d}$ is a scaling factor that aims to avoid too large or too small inner products that affect the stability of the gradient descent. The Softmax function calculates a weight for each Key vector, then multiplies these weights with the corresponding Value vector and sums the results to obtain the final output vector.

C. Equations

Finally, each output vector is stitched together to obtain the final output matrix Y, where the I^{th} (I=1…n) row represents the vector representation of the I^{th} (I=1…n) input element as in:

$$Y = [Attention(Q,K,V)_1, Attention(Q,K,V)_2, ..., Attention(Q,K,V)_n] \tag{5}$$

The improved Transformer network structure is shown in Fig. 3.

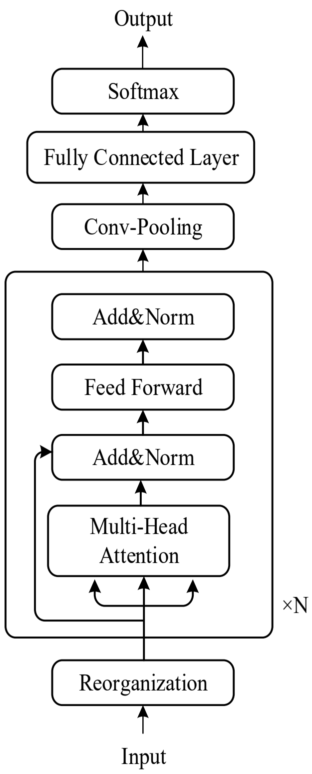

Fig. 3. Improved Transformer network structure.

The input sequence is encoded using the mechanism of Self-Attention to obtain a vector representation of each input element, and then these vectors are further processed by the encoder of the Transformer network to obtain the output results. Due to the sequential nature of flight chain data and there is also a relationship between each level of flight chain data, it is imperative to study the improved Transformer-based model for predicting the spread of flight delays. Fig. 4 shows the overall process of prediction.

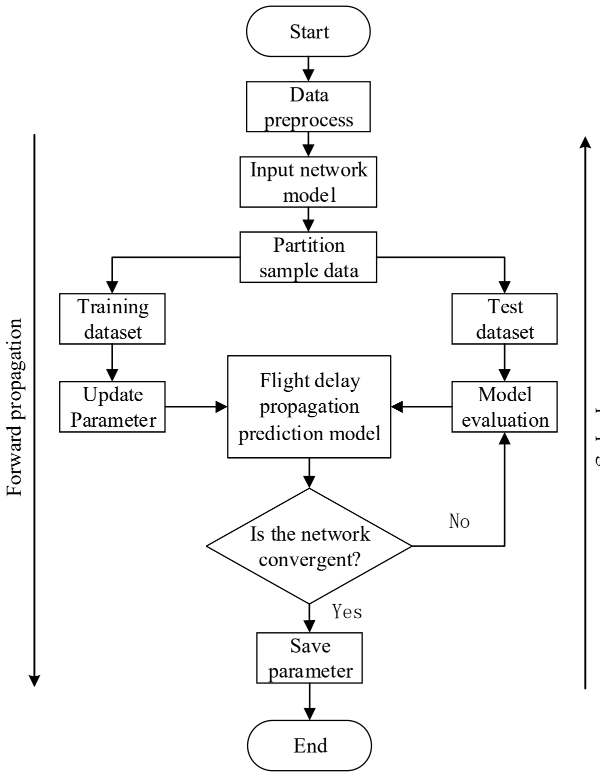

Fig. 4. Flow chart of overall flight delay propagation prediction model.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Comparative analysis of experimental indexes

To compare the effect of the models, the flight delay algorithm library models were introduced for the comparison of the experiments. The output result of the flight delay and prediction task in this paper is the flight delay level, which is a typical multi-classification task. In the multi-classification task, the main commonly used evaluation indicators include accuracy, percision, recall and F1 value, etc., which represent the model performance of different scales from different perspectives.

Accuracy is referred to the proportion of correctly classified samples in the total number of samples, reflecting the overall performance of the model algorithm. The calculation formula is as followed:

$$Accuracy = \frac{n_{correct}}{N} \tag{6}$$

Precision is referred to the proportion of all predicted positive results that are actually positive. Its formula is as followed:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

The recall is the ratio of the total number of real positive samples correctly predicted to the total number of real positive samples. The formula is as followed:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

F1 Score is the result of the comprehensive consideration, which is the harmonic average of accuracy and recall.

$$F1 = \frac{2 \times P \times R}{P + R} = \frac{TP}{TP + FN} \tag{9}$$

Where TP (True Positive) refers to the number of positive checks, FP (False Positive) refers to the number of false checks, and FN (False Negative) refers to the number of missed checks.

The introduced algorithms involving flight delay wave and prediction mainly include CBAM-CondenseNet-based flight delay wave and prediction algorithm model, SimAM-CNN-MLSTM-based flight delay wave and prediction algorithm model and Transformer-based flight delay wave and prediction model. In the training process of the neural networks, the models are trained using the training set, the parameters are adjusted to fit the data, and the training is examined using the test set to evaluate the generalization ability of the models. TABLE II. shows the specifics of loss values, accuracy, precision, recall and F1 values of the CBAM-CondenseNet model, SimAM-CNN-MLSTM model and Transformer model on the flight chain dataset, respectively.

TABLE II. ANALYSIS OF EXPERIMENTAL RESULTS.

Index	**CBAM-CondenseNet**	**SimAM-CNN-MLSTM**	**Transformer**
Loss value	0.3	0.2	0.358
Accuracy	0.898	0.9136	0.903
Precision	0.913	0.825	0.855
Recall	0.892	0.874	0.933
F1	0.904	0.849	0.897

The experimental results demonstrate the performance of the three algorithms in terms of extreme sample classification and overall accuracy. Specifically, CBAM-CondenseNet performs well in the classification of extreme samples, and the algorithm is able to effectively improve the model's classification ability for extreme samples by introducing an attention mechanism and compressing the network structure. SimAM-CNN-MLSTM performs better in terms of overall accuracy. This algorithm can make full use of the temporal information of the input data by combining convolutional neural network and multilayer LSTM network, thus improving the classification accuracy of the model.

In contrast, the Transformer algorithm uses techniques such as self-attentive mechanism and residual linkage, which can effectively handle the relationship between input sequences and adaptively adjust according to different input lengths, and can achieve relatively good accuracy while taking advantage of the adaptive flight chain length, thus improving the generalization ability and classification accuracy of the model.

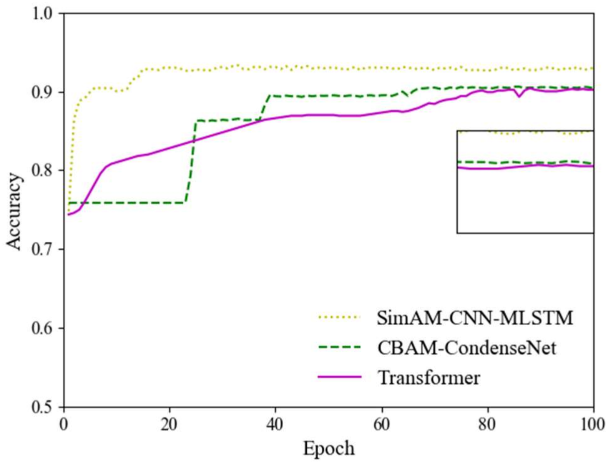

Fig. 5. Comparison of accuracy curves of different algorithms.

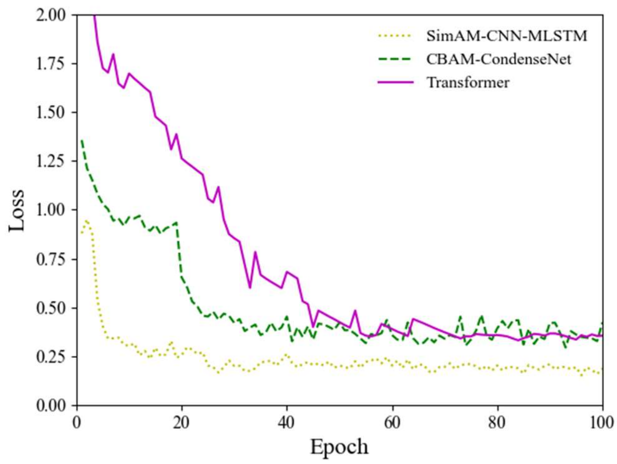

Fig. 6. Comparison of loss curves of different algorithms.

Based on the national flight data provided by the Administration of China Eastern Air Traffic Management, the accuracy change curves and loss value change curves of

SimAM-CNN-MLSTM model and CBAM-CondenseNet model and Transformer model during the training process are given in Fig. 5 and Fig. 6 respectively. From the trend of the curves, the SimAM-CNN-MLSTM model converges faster, and the accuracy rate is stabilized at about 91% after 100 rounds of model iterations, while the CBAM-CondenseNet algorithm and Transformer algorithm converge slower in the training process and finally stabilize at about 90%.

B. Algorithm Complexity Analysis

Time complexity and space complexity are two important concepts to describe algorithm efficiency. Time complexity refers to the measurement of the time required by the algorithm to solve the problem. It is often used to measure the execution efficiency of the algorithm, which is measured by FLOPs in this experiment. Spatial complexity refers to the measurement of the storage space required by an algorithm to solve a problem. It is often used to measure the memory occupancy of an algorithm, and the calculation is measured by the number of model parameter Params. The specific comparison is shown in TABLE II. .

TABLE III. COMPARISON OF MODEL COMPLEXITY.

Model	**FLOPs（M）**	**Params（M）**	**Prediction time（s）**
SimAM-CNN-MLSTM	4.83	1.05	5.42
CBAM-CondenseNet	20.77	1.46	5.79
Transformer	15.35	1.78	3.74

Usually, both time complexity and space complexity increase with the size of the problem. When designing algorithms, it is necessary to consider both time complexity and space complexity in order to achieve better efficiency and performance. As shown in the table, the Transformer model does not have much disadvantage in the complexity of algorithm evaluation, rather, due to its network structure has the advantage of better parallelization, Transformer network model has the fastest inference speed and better real-time performance, which can be faster to make predictions and decisions, and better meet the requirements of practical applications.

V. CONCLUSION

The issue of forecast flight delay propagation is studied in this thesis. The method based on Transformer is explored to pretreat the impact of flight delay. A dataset of triplet flight chain structure is constructed to effectively improve the accuracy of flight delay prediction. By inputting relevant data of flights, the delay degree of each level can be predicted, and reasonable slack time can be configured according to the predicted delay degree, which can absorb the delay impact of upstream flights and alleviate the situation of flight delay. This provides theoretical support for optimizing flight plans and controlling flight delays for relevant departments. According to the characteristics of flight delay, the Transformer model was adjusted and improved to enhance the accuracy and efficiency of prediction. The architecture of the network is restructured by inserting new convolutional pooling modules into the original network and adjusting the input and output dimensions of the network to make the overall model more tailored to flight delay prediction. Multiple aspects such as model design, parameter tuning, and experimental verification are involved to ensure the performance and robustness of the model. Finally, the results of the experiments show that the modified model has low complexity, short time and good real-time performance. It makes prediction and decision faster, improves real-time performance, better meets the needs of practical applications with an accuracy of 90.3%.

ACKNOWLEDGMENT

This research is funded by the Tianjin Municipal Education Commission Scientific Research Program, grant number 2022ZD006.

REFERENCES

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

[2] Li J, Zeng W, Liu D. Airport flight delay prediction based on spatio-temporal long-short term memory network [C]. China Association for Science and Technology, Ministry of Transport, PRC, Chinese Academy of Engineering. Proceedings of 2019 World Transport Congress (Part 2), 2019:504-517.

[3] Cui W. Flight delay prediction based on long-short term memory and residual network [D]. Zhejiang: Hangzhou Normal University,2019.

[4] Qu J, Ye M, Cao L. Airport delay prediction method based on hybrid coding and long/short time memory network [J]. Signal Processing, 2019, 35(07):1160-1169.

[5] Tsoi A C, Tan S. Recurrent neural networks: A constructive algorithm, and its properties[J]. Neuro Computing, 1997, 15(34):309-326.

[6] Hochreiter S, Schmidhuber J. Long-short term memory [J]. Neural Computation, 1997, 9(08):1735-1780.

[7] Noboru T, Ryosuke K, Akihide S, et al. Prediction of delay due to air traffic control by machine learning[C]. Proceedings of the 17th AIAA Modeling and Simulation Technologies Conference, Washington, USA, IEEE, 2017:191-199.

[8] Wu R, Zhao Y, Qu J. A Prediction model based on CBAM-CondenseNet [J]. Journal of Electronics and Information Technology, 2021, 43(1): 187-195.

[9] Qu J, Zhang J, Zhao Y. CNN-MogrifierLSTM flight delay prediction based on SimAM [J]. Journal of Signal Processing, 2022,38(11):2412-2423.

[10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. arXiv 2017, arXiv:1706.03762.