Article

A new way of airline traffic prediction based on LSTM graph convolutional network

Firstname Lastname [1,†,‡], Firstname Lastname [1,‡] and Firstname Lastname [2,*]

[1] Affiliation 1; e-mail@e-mail.com
[2] Affiliation 2; e-mail@e-mail.com
* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials) +xx-xxxx-xxx-xxxx (F.L.)
† Current address: Affiliation 3
‡ These authors contributed equally to this work.

Version January 20, 2021 submitted to Journal Not Specified

Abstract: With the development of society and the improvement of people's material level, airplanes are becoming more and more a means of transportation for people to travel. If an airline can predict the passenger flow in advance, it can be used as an important decision-making basis for its flight route planning, crew scheduling planning and ticket price formulation in the process of management and operation. However, due to the high complexity of aviation network, the existing traffic prediction methods generally have the problem of low prediction accuracy. In order to overcome this problem, this paper makes full use of graph convolutional neural network and long - short memory network to construct a prediction system with short - term prediction ability. Specifically, this paper uses the graph convolutional neural network as a feature extraction tool to extract the key features of air traffic data, and solves the problem of long term and short term dependence between data through the long term memory network, and builds a high-precision air traffic prediction system based on it. Finally, we design a comparison experiment to compare the algorithm with the traditional algorithm. The comparison results show that the algorithm has obvious advantages in air flow prediction.

Keywords: Graph Convolutional Network; Long Short Term Memory Network; Flow; Airlines; Predict

1. Introduction

In recent years, as an important industry in national economic and social development and an advanced mode of transportation, the demand for civil aviation passenger transport has been growing rapidly along with the rapid development of national economy and the substantial increase of people's income. Air traffic has always been the focus of attention because it is related to the development of civil aviation industry, the profitability of airlines and the vital interests of passengers[1].

Air traffic forecast plays an important role in aviation revenue management theory. The forecast results can be used to dynamically adjust the ticket price, which can make the airline company achieve the maximum revenue. However, the pricing mechanism of each airline is complex, and the real-time ticket price is constantly changing under the influence of many factors, which has the characteristics of trend, randomness and volatility. Therefore, how to forecast the air passenger flow accurately and reasonably has become an important content of China's air transport management.

Related researches have been conducted on air passenger flow forecast at home and abroad. An integrated learning binary classification algorithm HAMLET is proposed based on Q-learning [2]. HAMLET uses rule learning, reinforcement learning and time series technology, and combines their results through superposition generalization to produce the final prediction result. The accuracy can reach 74.5%. This method still has a lot of room for improvement in traffic forecasting. Groves, William and Gini, Maria [3] use a regression model for different prediction targets. The author introduces a complex feature extraction and selection method to predict, and predict the traffic two months in advance. A pre-processing technique, the Marked Point Process (MPP), was proposed to reduce the size of the feature set by retaining only the significant changes in the flow sequence and the turnover rate [4].

However, all the above studies focus on the optimal purchase timing problem, and seldom discuss the performance of regression prediction. The forecast of air flow can be naturally transformed into a time series forecasting problem. According to the time before departure, Zhao-Jun, Gu and Shuang et. al [5] use classical time series model to simulate the flow. Although many classic time series models are based on the assumption that there is linear relationship between the past and the future value, but the time series model to forecast the traffic demand data must have the correlation, due to the passenger flow influenced by various factors changing, trend and the characteristics of randomness and volatility, will also affect air flow prediction.

Applying machine learning technology to solve nonlinear time series problems [6] is a better trend. Various algorithms for machine learning have been studied and deployed [7–10]. Jiang et al.[11] applied grey SVM combined with empirical mode decomposition (EMD) to high-speed railway passenger flow prediction. Xie et al. [12] used EMD to model passenger flow prediction of airport terminals by least squares support vector machine. Support vector machine is regarded as a classification tool, and support vector regression aims to identify and optimize the error range of regression. If the observed value falls within the boundary, the cost of error is calculated as zero. However, SVM algorithm is suitable for small sample learning, and if the data sample size is too large, the prediction results will be adversely affected. The application of neural network in long - term demand forecasting is very popular. Tsai et al. [13] use them for short-term demand forecasting. They compared two advanced formulas, MTUNN and Penn, with the general multilayer perceptron model. Wei and Chen [14] used hybrid neural network model to predict passenger flow in rapid transit systems. Using neural network model to make prediction, we need to turn all the characteristics of the problem into numbers, and all the reasoning into numerical calculation, the result is bound to lose information. And model training needs a large number of data samples to support.

However, the above method based on machine learning is based on the historical data of each aviation station to forecast, this method does not consider the impact of the ticket price and passenger flow of other stations on the passenger flow of the current station, so the prediction performance is poor. Graphic Convolutional Neural Network (GCN), as a kind of neural network that can extract unstructured data, has gained a lot of attention in solving the relationship between adjacent points [15–17].

In view of the low accuracy of air passenger flow prediction at present and the trend, randomness and volatility of air traffic affected by many factors, we built a graph convolution-long short-term memory model based on graph convolutional neural network and long short-term memory neural network. In this model, the feature mapping of the data set is carried out by using the graph convolutional neural network. Then the long and short memory model is used to process the matrix data set and realize the prediction of ticket price. This method takes into account the different influences of various factors on the ticket price, and also combines the characteristics of the trend and volatility of the ticket price. The experimental results show that the graph convolution-long and short memory model can be used to predict the air ticket price well.

The remaining part of this paper is organized as follows. The main contribution of this paper is descripted in Section 2. In Section 3, two comparative experiments are used to prove the effectiveness of the algorithm. And conclusions are drawn in Section 4.

2. Main result

In airlines, the spatial distribution of aircraft stations is a non-euclidean structure, that is, the number of stations around each station is uncertain, and even if two stations are adjacent, they may not actually communicate with each other, resulting in no spatial relationship between their traffic. Therefore, traditional convolutional neural network (CNN) cannot accurately obtain their spatial information. At this point, multiple sites can be abstracted into A graph (see Figure 1). Features are extracted from the original input data to obtain the result of feature mapping of multiple channels. The intercommunication relationship between each site is represented by adjacency matrix A.

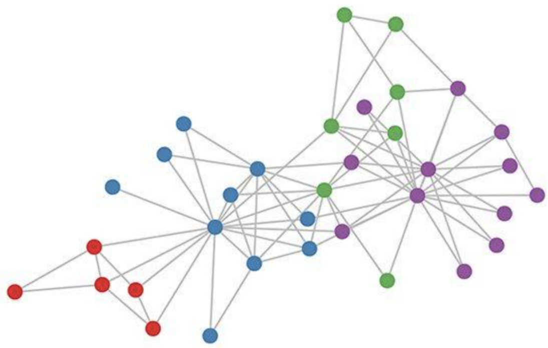

Fig. 1. The distribution of the stations.

2.1. The definition of the problem

The problem of airline passenger flow prediction can be described as follows: the historical flow data of each station $X_{t-s}, X_{t-s+1}, X_{t-2}, \cdots, X_{t-1}$ (s is the time step) can be used to predict the flow X_t of the next period. The formula is described as

$$X_t = F([X_{t-s}, X_{t-s+1}, \cdots, X_{t-2}, X_{t-1}]) \tag{1}$$

where, X is the site characteristics at each time step, and F is a nonlinear function.

In the actual traffic system, the network is regarded as a directed graph $G = (Q, V, A)$. Each sensor in the network is regarded as a node v_i and its value $Q \in R$ is a scalar. $V \in R^N$ and N is the number of sensors. The flow relationship between nodes consists of adjacency matrix A that is, the element A_{ij} in A represents the connection relationship between node V_i and V_j.

2.2. The description of the GCN

When dealing with the structure of the graph, it is necessary to obtain its Laplace matrix L, which is generally defined in the following ways:

$$L = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} = I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \tag{2}$$

where, $I_N$is the identity matrix of NN; Degree matrix D is defined as $D_{ii} = \sum_i A^i_{ij}$. Decompose the eigenvalue of L to get $L = \mathrm{L} = \mathrm{U}\Lambda\mathrm{U}^{\mathrm{T}}$. Λ is made up of L eigenvalues of diagonal matrix. $\mathrm{U} = \{\mathrm{u}_1, \mathrm{u}_2, \ldots, \mathrm{u}_\mathrm{N}\}$ is composed of the eigenvector L, and it is an orthonormal basis for R^N.

The spectral convolution theory in the graph structure has been supplemented and perfected in the paper. The convolution operation of convolution kernel G and input signal X in the time domain can be converted into the inner product form in the frequency domain.

$$g^*x = U\left(\left(U^{\mathrm{T}}g\right) \odot \left(U^{\mathrm{T}}\boldsymbol{x}\right)\right) = U_{g_\theta}(A)U^{\mathrm{T}}x \tag{3}$$

where $g_s(\Lambda) = U^T g = \mathrm{diag}(\theta), \theta \in R^N$, $\odot$ represents the hadamar product, $U^T g$means mapping g to the frequency domain space based on U. Due to g_θ high computational complexity, so using hierarchical linear model constraints and the chebyshev polynomial to approximate calculation. In this paper, The simplified first order polynomial form of $g * x$ is adopted.

$$g^* x = U_{g_\theta} U_x^{\mathrm{T}} \approx \theta \left(I_N + D^{-1/2} A D^{-1/2} \right) x \tag{4}$$

where $\widetilde{D}^{-1/2} \widetilde{A} \widetilde{D}^{-1/2} = I_N + D^{-1/2} A D^{-1/2} \quad \widetilde{A} = I_N + A$, $D = \sum_i \widetilde{A}_{ij}$, Therefore, the output of layer L is

$$H^{(l)} = \sigma \left(\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} H^{(l-1)} W^{(l)} \right. \tag{5}$$

where δ is the activation function, $\widetilde{W}^{(l)} = \theta^{(l-1)} W^{(l)}$, $\boldsymbol{\theta}^{(l-1)} \in \boldsymbol{R}^{c^{(1-l)} \times F^{(l-1)}}$, $\boldsymbol{W}^{(l)} \in \boldsymbol{R}^{F^{(l-1)} \times c^{(l)}}$, $C^{(L-1)}$ is the output dimension of the $(L-1)$ layer, and $F^{(L-1)}$ is the characteristic vector size of each dimension. Therefore

$$H^{(l)} = \sigma \left(\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} H^{(l-1)} \widetilde{W}^{(l)} \right) \tag{6}$$

At present, there is no effective measurement method for the calculation of adjacency matrix A. Most scholars use heuristic methods, that is, based on the Euclidean distance or Markov distance between sensors to determine the element value corresponding to the adjacency matrix. However, these methods all require manual calculation of the distance relationship between the sensors in advance. In this paper, the data-driven method is adopted to calculate the adjacency matrix, and $A = D$, then the formula can be written as:

$$H^{(l)} = \sigma / \widetilde{A} H^{(l-1)} \widetilde{W}^{(l)} \Big) \tag{7}$$

The element value of matrix A is learned from the sample data, that is, the matrix is composed of trainable parameters. The data-driven approach is more realistic than the heuristic approach. Therefore, the L layer of the convolutional neural network is constructed in accordance with Formula 7. It should be noted that the initial matrix A is the same for each layer of the convolutional network, and the parameters are updated only when the error is propagated backwards.

GCN introduces the spatial features of the graph by convolving the Laplace matrix with the input. In this paper, the model takes flight segments as nodes and the association between flight segments as edges to build a graph. According to the graph, the adjacency matrix is obtained and the demand of future flight segments is predicted by combining the price and demand of historical flight segments.

2.3. *The problem of time series*

It is found that Recurrent Neural networks (RNN) are widely used in sequential data such as natural language and image processing, which have a significant effect. Since then, various types of circulating neural networks have been proposed. Aiming at the problem of air passenger flow prediction, this paper introduces the Long Short-Term Memory network (LSTM), which can extract the characteristic information of the input sequence and find its internal relation, so as to improve the prediction accuracy of the model. In order to make use of the spatial and temporal characteristics of the data at the same time, the GCN model is combined with the LSTM model, and the GCN module is added to the output of the upper level.

The LSTM network structure used in this paper is shown as Figure 2.

The network model mainly accepts three inputs: X, H and C represent the current state, hidden layer state and cell state, respectively.

LSTM mainly realizes the management of long and short term memory through three gating units. The first step in the LSTM is to determine what information needs to be thrown out of the cell state. This decision is made by a sigmoid layer called the "Forget Gate." Input x and h, output a number

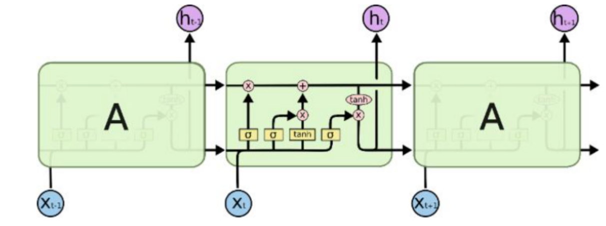

Fig. 2. The structure of LSTM.

between 0 and 1. The value of 1 means "keep the value completely", while 0 means "throw the value away completely". The formula of forgetting gate is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{8}$$

where W_f and b_f are the parameters to be learned, and σ is the sigmod activation function.

The second step is to determine what information we need to store in the Cell State. There are two parts to this question. First, a sigmod layer calls the "Input Gate" to determine which data needs to be updated. Then, a tanh layer creates a vector C_1.

$$\begin{aligned} i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \\ \tilde{C}_t &= \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \end{aligned} \tag{9}$$

After deciding what needs to be forgotten and what needs to be added, the old cell state C_{t-1} can be updated to the new cell state C_t.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{10}$$

Finally, we need to decide what to export. This output is based on our cell state, but will be a filtered part of the value. First, we run an output gate to determine which part of the cell state we are going to output. Then we put the cell state into the tanh (pressing the value between -1 and 1), and finally we multiply it by the output of the output gate.

$$\begin{aligned} o_t &= \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \\ h_t &= o_t * \tanh\left(C_t\right) \end{aligned} \tag{11}$$

2.4. *The description of algorithm*

The network structure based on GCN-LSTM model proposed in this paper is shown in the Fig. 3. The model mainly adopts encoder-decoder structure. In the encoder, multiple parallel GCN modules are used to extract the key features of the graph network with different time series. Then, the extracted time series features are transmitted to LSTM, and feature analysis and further feature extraction are carried out on the sequence data through LSTM to solve the long-term and short-term dependencies between the data. Finally, the encoder generates an encoded pair vector and sends it to the decoder. In the decoder, the multi-layer feedforward neural network is used to further process the features of the coding vector. Finally, the processed data is transmitted to a GCN network to produce predicted values.

In order to improve the prediction performance of the GCN-LSTM algorithm, in this paper, we use the final output through the network and the value of the real label to calculate L_1 loss, also known as the mean absolute error (MAE), and take the mean square error (MSE) as the evaluation index of the model. The specific calculation formula is as:

$$\text{loss} = MAE = \frac{1}{m}\sum_{i=1}^{m} |(y_i - \hat{y}_i)| \tag{12}$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m} ((y_i - \hat{y}_i))^2 \tag{13}$$

In this paper, the Adam-based batch gradient descent optimization algorithm is used to learn and update parameters through the loss function to minimize the loss function until the loss function converges. The trained model is used to predict the test set and calculate the MSE. The smaller the MSE value is, the closer the predicted value of the model is to the real value and the better the generalization ability of the model is.

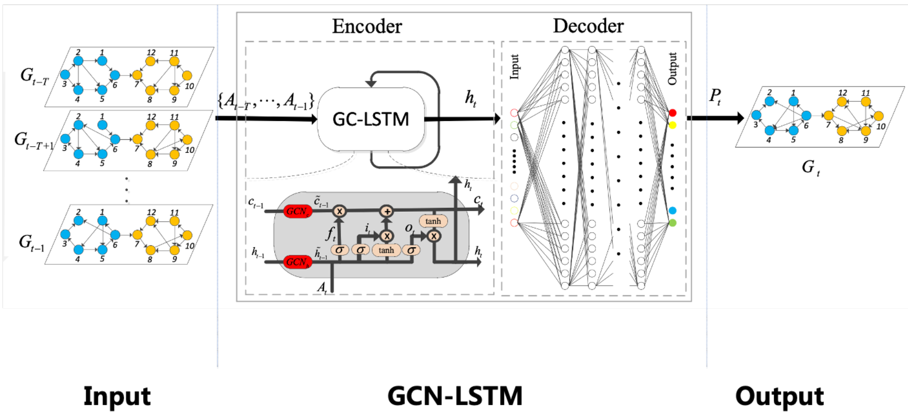

Fig. 3. Overall structure of GCN-LSTM model (original features of air passenger flow data are extracted by using first-order approximate GCN, and output features are analyzed by LSTM for long-term and short-term sequence characteristics, and then predicted values are obtained).

3. Experiments

This paper selects 6,9111,332 ticket sales data examples of 28,809 flights of 17 domestic airlines between September 1, 2020 and October 31, 2020 to verify the feasibility and effectiveness of the GCN-LSTM combined prediction model. The selected data set is relatively complete without missing values. After obtaining a complete data set, an adjacency matrix is constructed according to the data set. Each flight segment is a node. Between nodes, if the origin is the same, it is considered to have a diverting effect on traffic, and the weight is set as the number of (0,1). If the destination is the same and the traffic is considered to be promoted, the weight is set as the number greater than 1 and the rest as 0.

Five sections of AAT_URC, CAN_PEK, CAN_CSX, CAN_CTU and CAN_CKG are taken as examples to construct an adjacency matrix, in which AAT, URC, CAN, PEK, CSX, CTU and CKG are city names.

Table 1. Adjacent Matrix.

	AAT_URC	CAN_PEK	CAN_CSX	CAN_CTU	CAN_CKG
AAT_URC	0	0.0	0.0	0.0	0.0
CAN_PEK	0	0.0	0.5	0.5	0.5
CAN_CSX	0	0.5	0.0	0.5	0.5
CAN_CTU	0	0.5	0.5	0.0	0.5
CAN_CKG	0	0.5	0.5	0.5	0.0

Time series features are constructed. The above five flight sections are also taken as examples. According to the take-off time, the price series 14 days before the take-off time is taken as the X feature. The flow of 1 day before the take-off was taken as Y, and such input and output were a sample. The take-off time was calculated forward to increase the number of samples. In this way, there are 14 days' time data, 1 piece of demand data, arranged in a time series according to time.

In this paper, the data set of two months is divided into training set and test set according to the ratio of 8:2, normalized, and trained by GCN-LSTM. The timing step size of the model is 15, and the window is constantly moved to predict. Under the premise of the same input of historical price time series, it is compared with the experimental results of traditional AR model, Moving Average, Exponential Smoothing, LSTM and SVR.

Table 2 lists five precise results of route prediction, which show that the GCN-LSTM model can improve the accuracy of prediction.

Table 2. Prediction results MSE.

	GCN-LSTM	LSTM	SVR	AR	MA	ES
Average of five routes	6.86%	10.84%	12.72%	14.46%	13.73%	16.93%
AAT_URC	5.44%	8.71%	10.73%	11.37%	10.74%	14.39%
CAN_PEK	6.91%	9.74%	12.02%	12.75%	11.71%	16.47%
CAN_CSX	8.87%	12.03%	13.07%	17.13%	14.57%	18.29%
CAN_CTU	6.24%	11.73%	14.35%	15.01%	13.79%	17.48
CAN_CKG	6.84%	11.99%	13.43%	16.04%	17.84%	18.02%

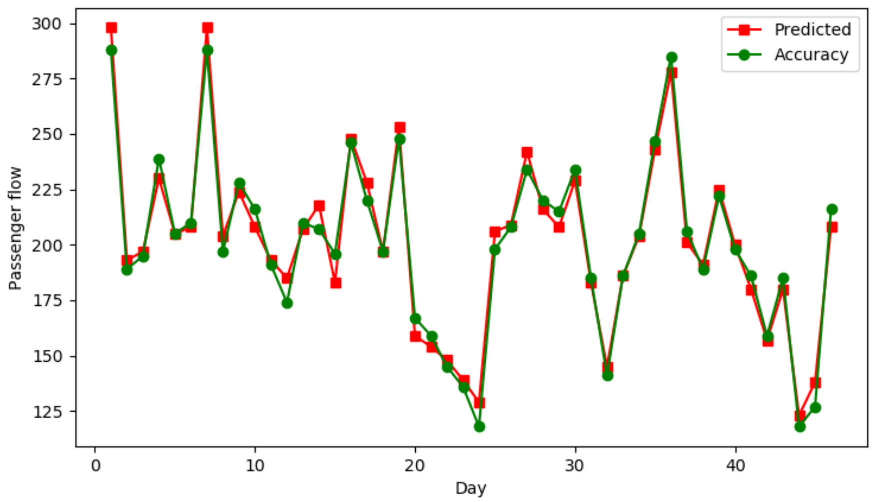

Fig. 4. OGCN-LSTM model prediction results of AAT URC route.

Fig.5 shows the average MSE values predicted by AR model, moving average, exponential smoothing, LSTM and SSR for five airlines. The red line is the prediction results of GCN-LSTM model.It can be seen from the figure that the GCN-LSTM model is used for prediction, and the error fluctuation is small, and the MSE is basically in the range of 5%-9%.The prediction results of this model are obviously better than those of other models.

4. Conclusion

In view of the problems existing in air flow prediction, an air flow prediction model based on graph convolutional neural network and long short-term memory network is proposed based on in-depth analysis of the influencing factors of air flow prediction. Firstly, based on the characteristics of air traffic data, a feature extraction network based on graph convolutional neural network is designed. Then combined with the long - short memory network to solve the problem of long - short - term data

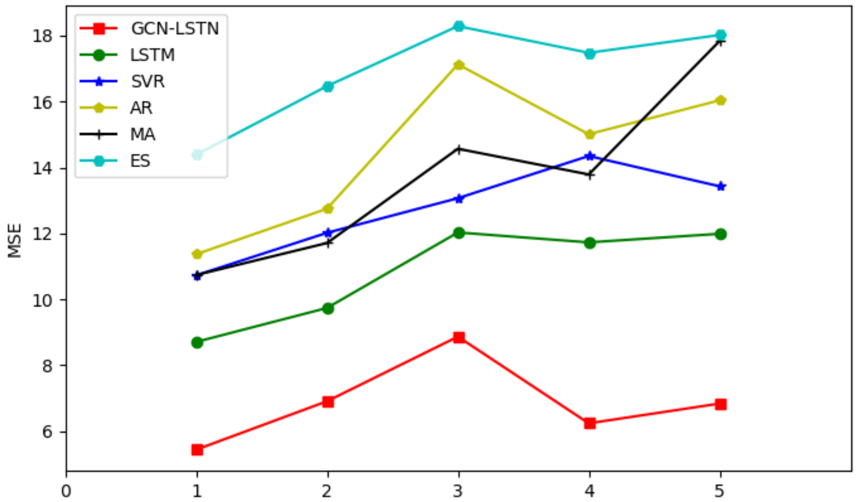

Fig. 5. Comparison of MSE results of five airlines.

dependence; Finally, the prediction results are outputted based on the feedforward neural network. In the experimental part, we verify the performance of the GCNLSTM model on aviation data sets. The experimental results show that the prediction results of this model are obviously more accurate than the existing algorithms, and it has higher prediction performance.

References

1. Donovan, A.W. Yield management in the airline industry. *Journal of Aviation/Aerospace Education & Research* **2005**, *14*, 9.
2. Etzioni, O.; Tuchinda, R.; Knoblock, C.A.; Yates, A. To buy or not to buy: mining airfare data to minimize ticket purchase price. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 119–128.
3. Groves, W.; Gini, M. A regression model for predicting optimal purchase timing for airline tickets. *University of Minnesota, Minneapolis, MN, Tech. Rep* **2011**, pp. 1041–4347.
4. Wohlfarth, T.; Clémençon, S.; Roueff, F.; Casellato, X. A data-mining approach to travel price forecasting. 2011 10th International Conference on Machine Learning and Applications and Workshops. IEEE, 2011, Vol. 1, pp. 84–89.
5. Zhao-Jun, G.; Shuang, W.; Yi, Z. Flight ticket fare prediction model based on time-serial. *Journal of Civil Aviation University of China* **2013**, *31*, 80.
6. Bontempi, G.; Taieb, S.B.; Le Borgne, Y.A. Machine learning strategies for time series forecasting. European business intelligence summer school. Springer, 2012, pp. 62–77.
7. Carmona-Benítez, R.B.; Nieto, M.R. SARIMA damp trend grey forecasting model for airline industry. *Journal of Air Transport Management* **2020**, *82*, 101736.
8. Nieto, M.R.; Carmona-BenÃtez, R.B.; others. ARIMA+ GARCH+ Bootstrap forecasting method applied to the airline industry. *Journal of Air Transport Management* **2018**, *71*, 1–8.
9. Madhavan, M.; Ali Sharafuddin, M.; Piboonrungroj, P.; Yang, C.C. Short-term Forecasting for Airline Industry: The Case of Indian Air Passenger and Air Cargo. *Global Business Review* **2020**, p. 0972150920923316.
10. Pan, B.; Yuan, D.; Sun, W.; Liang, C.; Li, D. A Novel LSTM-Based Daily Airline Demand Forecasting Method Using Vertical and Horizontal Time Series. Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 2018, pp. 168–173.
11. Jiang, X.; Chen, X.; Zhang, L.; Zhang, R. Dynamic demand forecasting and ticket assignment for high-speed rail revenue management in China. *Transportation Research Record* **2015**, *2475*, 37–45.

12. Xie, G.; Wang, S.; Lai, K.K. Air passenger forecasting by using a hybrid seasonal decomposition and least squares support vector regression approach. Proceedings of the 59th world statistics congress of the international statistical institute, Hong Kong, 2013.
13. Tsai, T.H.; Lee, C.K.; Wei, C.H. Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Systems with Applications* **2009**, *36*, 3728–3736.
14. Wei, Y.; Chen, M.C. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies* **2012**, *21*, 148–162.
15. Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; Liu, Y. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, Vol. 33, pp. 3656–3663.
16. He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; Wang, M. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. *arXiv preprint arXiv:2002.02126* **2020**.
17. Wei, Y.; Wang, X.; Nie, L.; He, X.; Hong, R.; Chua, T.S. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1437–1445.

Sample Availability: Samples of the compounds ...... are available from the authors.