

Short-term Forecasting Method of Air Traffic Flow based Neural Network Ensemble

Ming Zhang, Ming Zhang, Kai Liu, Hui Yu, Jue Yu

(School of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016)

Keywords: air transportation; short-term prediction; disuniform data; neural network ensemble; k-means clustering; 3σ Principle; fuzzy membership

Abstract: In this research, in order to address interferences of air traffic from complex factors like weather and local data abnormality of radar samples, fuzzy clustering and neural network ensemble were introduced into the short-term forecasting of air traffic flow. Firstly, with K-means cluster analysis, this research compared traffic volume at different time with that of each clustering center to identify the temporal clustering of traffic volume. Secondly, according to different data sets from clustering analysis, corresponding neural network models were established. On the basis of Bagging method, a neural network ensemble weight allocation algorithm of fuzzy subordinative degree was also built to identify weight of each neural network and to establish neural network ensembles model. Finally, according to 3σ principle of normal distribution, abnormal data out of Section ($\mu - 3\sigma, \mu + 3\sigma$) was cleaned and short-term forecasting results were acquired. Our model showed superior results of short-term radar data forecasting for Shanghai Terminal Area, overmatching regression analysis and neural network forecasting. The experiment verified that the method is valid and feasible for short-term forecasting of air traffic flow.

1. Introduction

In recent years, with the rapid development of civil aviation transportation industry, conflicts between airspace resources and air traffic demand have become increasingly serious. Accurate forecasting of air traffic flow in terminal airspaces is a premise and foundation for the terminal airspace resource optimization and air traffic flow management, thus showing important research value. However, as short-term forecasting of air traffic flow should take into account many influential factors, it is accompanied with complexity and randomness, which are obstacles for accurate forecasting. This is considered as a bottleneck for air traffic flow management and airspace management research.

To this end, researchers all over the globe are engaging in the research of this area and have reached certain progresses. For example, Faraway J et al ^[1] followed the characteristic that neural network can memorize complex patterns and trends of historical data, carried out the learning of a great amount of historical flight traffic data, and established an air traffic flow forecasting model based on neural network. Ilona W et al ^[2], on the basis of historical data, identified parameters of time series model and established a time series model for civil aviation traffic forecasting. Cui Deguang et al ^[3-5] analyzed influence factors of civil aviation traffic flow, and proposed an artificial neural network combined with regression method for the traffic flow forecasting of Dawangzhuang navigation station. They acquired forecasting results superior to artificial neural network. Liu Yumei et al ^[6] applied system identification theory to identify model and parameters and established an air traffic flow forecasting model based on the principle of least square estimation. Xu Xiaohao et al ^[7] proposed a combined forecasting method based on different data sources. Through the statistics and forecasting of flight time, flight planning data and radar data, they used genetic algorithms for model weight optimization and established a combined forecasting model of minimum error. Zhao Yifei et al ^[8] proposed a Holt-Winters model based on “optimal smoothing coefficient” to forecast short-term air traffic flow of airport. However, the above findings mainly focused on the research of forecasting methods, and attempted to improve the precision of forecasting by selecting and combination of forecasting methods. Yet, it is known that influencing factors of precision of forecasting include forecasting methods and forecasting sample data. The latter is the source and

basis of the entire forecasting. Currently, radar surveillance control is now widely adopted by major terminal areas in China. However, due to the influences from complex random factors such as the weather, the economy and fuel prices [9], radar forecasting data present local abnormality of data, which is a key reason for the difficulty of improving precision of forecasting.

Because of uncertainties of air traffic in terminal areas, the application of regression method in forecasting is limited.

However, the time series analysis methods are more in line with the characteristics of flow series, especially for systems that are hard to establish a precise mathematical model with a variety of uncertainties and non-linearity. Artificial neural network and other intelligent forecasting methods can often handle problems that are hard to be addressed by traditional methods. However, related theories and practices have shown that [10, 11] pure neural network forecasting model presented a defect that with the increase of complexity, more hidden nodes will be generated and the training time will be extended greatly, resulting in training difficulties. To solve the defect, infinite data is needed to generate perfect effects. In fact, training data is generally very limited. A single network will be difficult to adequately reflect and mine information, tending to make unsatisfactory training precision. Excessive pursue of training precision often leads to “overly fitting”. Meanwhile, neural network ensemble can dig information of several single networks to improve neural network performance. Therefore, this research proposed a short-term forecasting method of terminal area based on neural network ensemble. Through the establishment of intelligent data clustering and self-adapting neural network ensemble forecasting model, this research cleaned local data of abnormality and acquired precise short-term forecasting results of traffic volume.

2, modeling of fuzzy neural network ensemble

In short-term forecasting, when short-term forecasted traffic volume shows substantial variation in each time period, the data should be classified according to traffic volume so as to obtain several sets of data with relatively consistent traffic volume. Therefore, intelligent classification of data is a precondition and foundation for the forecasting with neural network ensembles model. This research adopted K-means cluster analysis to realize the classification of air traffic flows.

2.1 Clustering analysis to enable the classification of short-term forecasting of traffic volume

For radar data of air traffic flow, the clustering technology was used to divide the input space into several sub-space with more obvious features (eg. different traffic periods), and for different sub-spaces, corresponding neural network models with different structures were established. The forecasting was then conducted by neural network ensemble. Our research adopted algorithm to determine the clustering center, and compared traffic volume at different times with traffic volume in each clustering center to determine the temporal clustering and division of traffic volume. The specific algorithm is as follows:

A data set contained n traffic data was given and the category k was given by users. k data objects were randomly selected as the initial clustering centers. The remaining data objects were assigned to the most similar clusters according to the principle of greatest similarity. Next, according to the existing clustering results, center of mass in existing clusters was used to define the prototype. k centers of mass were taken as new clustering centers for re-clustering of data sets. The process was repeated until all the defined criterion functions tended to a given value or began to converge. The criterion function for the sum of squared errors is defined as:

$$J_e = \sum_{i=1}^k \sum_{p \in x_i} \|p - m_i\|^2 \quad (1)$$

Where, J_e is the sum of squared errors, p is spatial data object, m_i is the mean of Cluster x_i . The process was as follows:

- (1) According to the investigation of air traffic of target airspace, the time series distribution of short-term forecasting of traffic volume was identified and k objects were determined.
- (2) Each object was taken as the “center” of a category, representing k -th category.
- (3) According to the principle of nearest to the center, most similar clusters to each object were

- sought and other objects were assigned to each corresponding category.
- (4) For each category, the mean of all objects was calculated as the new “center”.
 - (5) According to the principle of nearest to the center, all objects were re-assigned to each category.
 - (6) Back to (3) until no changes of objective function were observed.

According to the above process, k clustering centers were obtained. With the air traffic flow $x_i(i=1, 2, \dots, k)$ as the reference, air traffic flow in each time period was divided.

2.2 Air Traffic Neural Network Ensembles Model

RBF network can realize the non-linear conversion from input space R^N to output space R^M through the linear combination of primary function. Meanwhile, air traffic flow data is a type of time series with relatively intensive non-linearity. This method can forecast M data in the future from the previous N data. In essence, this method is to identify a non-linearity mapping relationship from R^N to R^M . therefore, RBF network is appropriate for the forecasting of systems containing non-linear time series, such as the short-term forecasting of traffic volume.

For air traffic flow forecasting, some scholars have proposed a forecasting method based on neural network [3, 4], which is proven to be an efficient non-model method. However, the configuration and training of neural network is a NP issue. In the practical application, due to the lack of a priori knowledge of the problem, a great amount of laborious time-consuming experiments are needed to identify proper neural network models, algorithms and parameter settings. The effects of application are completely determined by users’ experiences, which influence the improvement of the network generalization. Hansen and Salamon^[10] proposed a pioneering neural network ensemble, which can significantly improve the network generalization by training several neural networks and the integration of results. Neural network ensembles mainly encompass Bagging and Boosting. Bagging shows certain differences from Boosting. The training sets of Bagging are randomly selected and each set is independent from the other. Meanwhile, the selection of Boosting training sets is not independent. The selection of each set is correlated with the previous round of learning result. Moreover, each forecasting function of Bagging does not have weight, while Boosting does. In addition, each forecasting function of Bagging can be generated in a parallel manner but that of Boosting can only be generated in order. Compared with time-consuming learning methods like neural network, Bagging can save a lot of time by parallel training.

Therefore, the application of a pure method in air traffic forecasting has certain forecasting errors. This is mainly because that in a period of time, air traffic in the terminal area may show significant variation (such as traffic volume of daytime and nighttime, traffic volume during holidays and regular days). Besides, changes of short-term forecasted traffic volume are not mutation but present decreasing or increasing progressively^[12]. To this end, this research obtained different spaces of data sets based on clustering analysis, and established corresponding neural network models for different data sets. With Bagging, a weight allocation algorithm for neural network ensemble of fuzzy sets was established to identify the weight of each neural network. Finally, the results of neural network ensemble were obtained. The method of neural network ensemble is shown in the following figure.

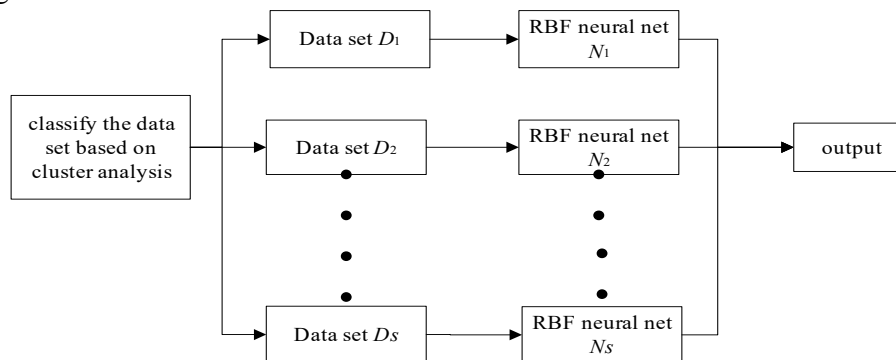


Figure 1 Diagram of neural network ensemble

The weight allocation algorithm of neural network ensemble is designed as follows: assume the above clustering method divided s traffic periods of a day, and s sample spaces were divided

according to respective time period. The training set of each period was the ensemble result by training an individual neural network and weighing the network. Given the traffic transitivity among time periods, the weight of each neural network was calculated by the fuzzy subordinative degree at forecasting points in the time period. We assumed that the clustering results were divided into two groups: daytime and nighttime. The fuzzy curves are shown in the following figures:

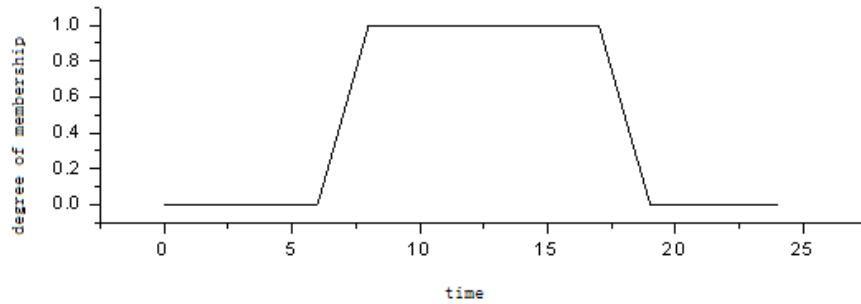


Figure 2 Fuzzy Curve of Traffic Flow of Daytime

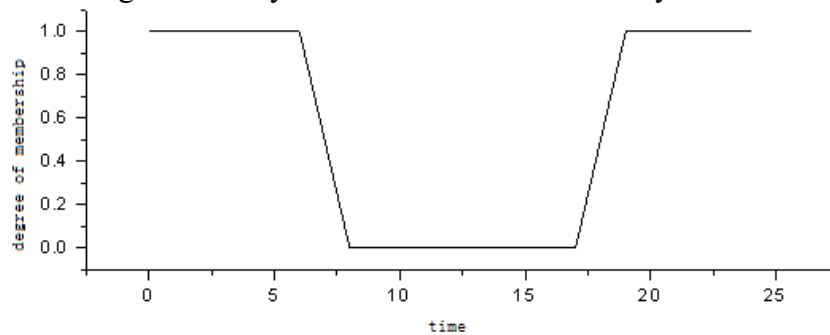


Figure 3 Fuzzy Curve of Traffic Flow of Nighttime

Let the forecasting point be t . Its membership to rush hours of daytime, other hours of daytime, and nighttime were $f_1(t)$ and $f_2(t)$ respectively. Therefore, the corresponding weight of neural network output in each time period is:

$$k_i = \frac{f_i(t)}{f_1(t) + f_2(t)} \quad i=1, 2 \quad (2)$$

When applying the method broadly, if s sets of sample spaces are obtained by clustering analysis, the corresponding weight of neural network output in each time period is:

$$k_i = \frac{f_i(t)}{\sum_{i=1}^s f_i(t)} \quad i=1, 2, \dots, s \quad (3)$$

This method ensures the precision of forecasting, reduces the size of each network, and shortens the learning time. Therefore, it enables a sound ability to learn.

2.3 Abnormal data cleaning and forecasting based on neural network ensembles model

There are many methods for abnormal data detection [13, 14]. The conventional outlier detection is based on whether the current observed value exceeds the pre-established threshold. Among methods of the threshold detection, generally we set a normal parameter base line according to the historical data, that is, a permissible range (containing 2 or 3 standard deviations). Once the value exceeds the range, it is identified as an outlier. However, the method also shows some problems, such as how to set a proper threshold, and hard to identify some subtle abnormal behaviors in the short-term forecasting of traffic volume. Besides, as flows in air traffic network have significant differences in different time periods, one threshold for all time periods is not accurate.

Our research considered that outliers do not solely indicate that an observed value exceeded the set threshold. Instead, it should also include non-conformities from the rules of data distribution in a period. Therefore, the pure dependence on the threshold for outlier identification often leads to a failure of identifying all outliers. However, from the perspective of forecasting error in neural network ensemble, it is easy to detect all outliers. This is because neural network can capture the

basic distribution patterns of traffic data while abnormal sequences are not consistent with basic rules. Therefore, neural network presents obvious forecasting errors. For the short-term forecasting of air traffic, due to the randomness of short-term forecasting of traffic volume, abnormality samples could be observed among some samples. These abnormality samples inevitably cause jeopardized precision of forecasting. However, due to the excellent generalization ability, neural network ensembles model can enable the fitting of most air traffic flow data in terminal areas. Besides, it supports the identification of a small amount of abnormal data, which can remarkably improve forecasting results. Hence, if several neural networks present large forecasting errors of the same part of data, this part of data can be identified as abnormal sequences.

With the stochastic oscillation of time series, each forecasting error is a random variable, and each random variable of all samples is mutually independent and distributed. According to the central limit law of probability theory, the approximation of forecasting deviation is subject to rules of normal distribution (when the number of random variables tends to infinite, it is subject to normal distribution). 3σ criterion is based on repeated precision measurements like normal distribution, and the interference or noise caused singular data cannot meet the requirement of normal distribution. Therefore, according to 3σ criterion of normal distribution, abnormal data of small probability events outside the Section $(\mu - 3\sigma, \mu + 3\sigma)$ were cleaned.

The cleaning method of neural network ensemble for abnormal data is as follows:

(1) $e_i(x)$ represents the forecasting error of i -th neural network for Sample x . Assume that forecasting error is subject to normal distribution, i.e. $e_i(x) - N(\mu_i, \sigma_i^2)$. The mean μ_i and variance σ_i^2 of forecasting error can be calculated according to the forecasting errors of the forecasting model in the training set D .

(2) According to 3σ criterion of normal distribution, for a single forecasting model, the identification function is:

$$\delta_i(x) = \begin{cases} 1 & \text{if } |e_i(x) - \mu_i| \leq 3\sigma \\ 0 & \text{else} \end{cases}, \quad (4)$$

From the above equation, it is known that $\delta_i(x)=0$ represents the expectancy value of i -th forecasting model for a forecasting error at forecasting point x with great deviation.

(3) When the above fuzzy subordinative degree is used to identify neural network ensemble, the majority of the voting in neural network ensembles model is used to identify whether a sample is an outlier, that is:

$$\bar{\delta}(x) = \frac{1}{s} \sum_{i=1}^s \delta_i(x) \quad (5)$$

where, the number of integrated neural networks is s , or a number of clusters. The forecasting errors of Sample x satisfied $\bar{\delta}(x) \geq \gamma$, $0.5 < \gamma \leq 1$, and smaller than the threshold γ . Thus, it is identified that the Sample x was an outlier. The value of the threshold γ was subject to the amount of forecasting data.

3 Illustrative Example of Short-Term Forecasting of Traffic Volume

3.1 Clustering analysis of flight data recorded by radar

With radar recorder, the statistical radar data of Wuxi corridor of Shanghai Terminal Area were used to obtain the traffic flow data in two time periods (from May 3, 2013 to May 12, 2013, and from October 11, 2013 to December 24, 2013). With the Traffic statistics in each two hours per day, after 85 days of data collection, a total of 1020 samples of data were acquired. Traffic statistics results showed that the short-term forecasting load of traffic volume of daytime greatly exceeded that during the night. Therefore, daily traffic volume in different time periods was considered to be divided into two parts. SPSS was used for clustering analysis of data, with the cluster number of 2 and the probability of distance differences between categories smaller than 0.001. The original 12 clusters were aggregated into two categories. The first included two periods: Period 1-4 and Period 10-12. The other category encompassed Part 5-9. In other words, data sets from 0:00 to 8:00 and

from 18:00 to 24:00 were taken as one category. Data sets from 8:00 to 18:00 were taken as other category of data sets.

In accordance with the above clustering results, after identifying two categories of data sets, this research identified neural networks for different data sets, and each network weight of neural network ensemble. According to Figure 2 and 3 in Section 1.2 as well as Eq.2, neural network weights of two categories of data sets were identified.

RBF neural networks of two categories of data sets determined by the above clustering analysis were respectively established. 75 groups of 85 groups of data were taken as training samples and the other 10 groups were testing samples. Structure of RBF neural network of daytime data sets was 5-6-1, and that of nighttime was 7-7-1. The distribution density of SPREAD radical primary function of two neural networks was determined by trials.

3.2 Data Cleaning and Forecasting Based On Neural Network Ensembles Model

(1) Data Cleaning of Neural Network Ensembles Model

According to abnormal data cleaning steps given in Section 1.3, MATLAB was used to obtain neural network forecasting results of the two data sets. For $ei(x)$ of forecasting result (forecasting error of i -th neural network to Sample x), MATLAB curve fitting function was used to identify that for normal distribution $ei(x) - N(i, i2)$, mean forecasting error I and variance I were 6.06 and 2.6 respectively. Finally, according to Eq.4 and 5, outliers of forecasting errors of 85 groups of samples that did not meet the threshold values were detected (the threshold γ was taken as 0.6).

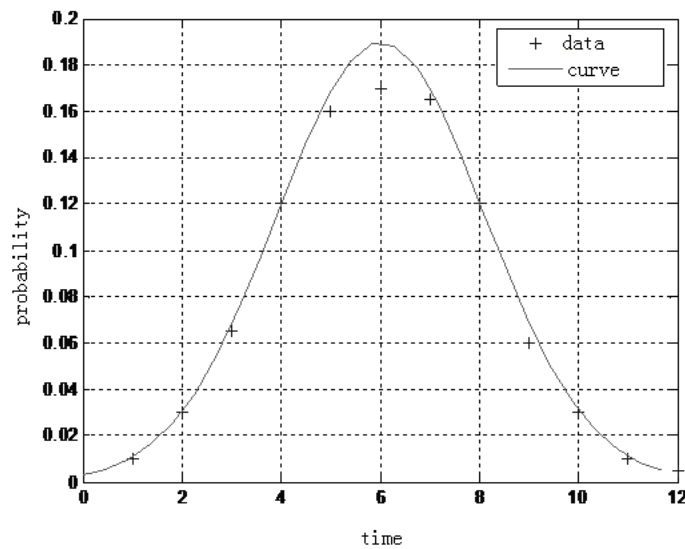


Figure 4 Normal Distribution Curve Fitting of Forecasting Error

According to the forecasting of 85 groups of data, the forecasting error $ei(x)$ was acquired. For each point of forecasting error, Origin was used for normal distribution curve fitting and identified a forecasting error (outlier) between 10:00 to 12:00 a.m. on May 3, 2012. Neural network ensembles model then enforced forecasting of data sets after the elimination of outliers.

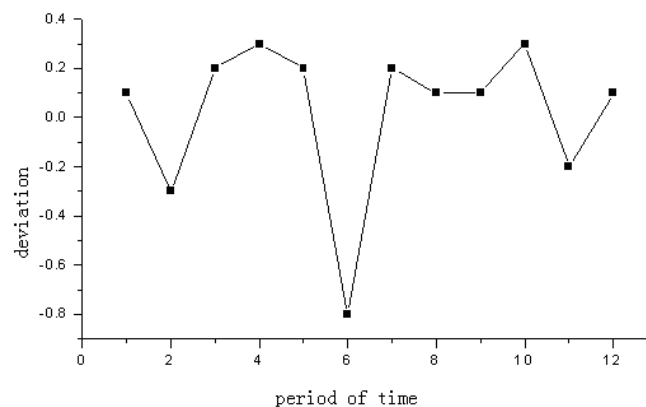


Figure 5 Identify errors in each time period according to neural network

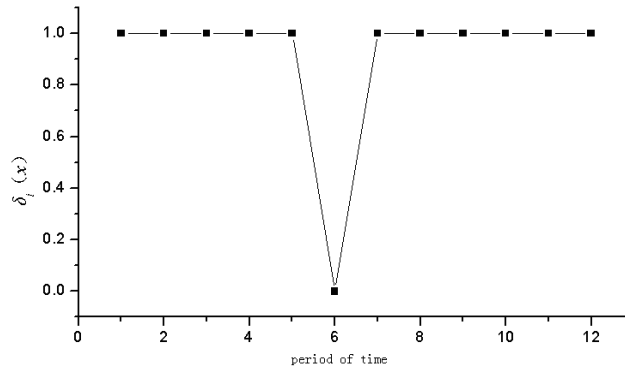


Figure 6 Identify outliers according to 3σ criterion of normal distribution

(2) Neural Network Ensemble Forecasting Based On Data Cleaning

This research used MATLAB to determine the fitting curve of normal distribution. After the data cleaning of outliers that did not meet the threshold, the sample data sets and neural network ensembles model without outliers can be used to generate short-term forecasting results.

The comparison of relative difference in forecasting results between neural network ensemble (NN ensemble), neural network (NN) and regression analysis (RA) is shown in the following figure. Among the three methods, neural network model was a RBF neural network model, with a structure of 12-14-1. The distribution density of SPREAD radical primary function was 3. 75 groups of 85 groups of data were taken as training samples and the other 10 groups were testing samples. MATLAB (RBF neural network forecasting) and SPSS (regression analysis) were used for forecasting and acquiring forecasting errors. For forecasting errors of the three methods, the relative difference was identified for comparison and the comparison results are shown in the following figure.

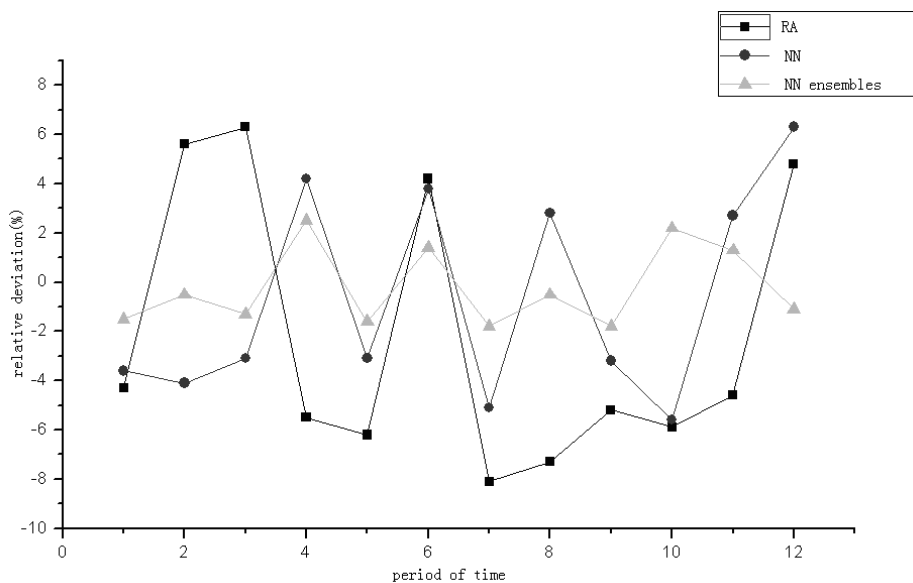


Figure 7 A Comparison of relative difference between neural network ensemble forecasting model, regression analysis and RBF neural network model

With the above method, daily changes of flight traffic volume in each sector and terminal airspace of Shanghai terminal airspace could be processed into statistics and forecasted, thus generating short-term traffic flow forecasting of the terminal airspace.

4 Conclusion

This research proposed to integrate sample difference analysis into self-adapting neural network ensembles model, in order to accomplish abnormal data identification. The data processing of abnormal forecasting samples replaced the previously subjective “cleansing abnormal data” during the data preparation phase. Instead, this paper adopted clustering analysis and self-adapting neural

network ensembles model to intellectually identify abnormal data, and acquired relatively more accurate forecasting results. As short-term forecasting of air traffic is also influenced by factors like temporary flights, military aircraft activities and traffic control, the mutability and randomness are significant. An in-depth study on mechanisms of flight operations and air traffic flow management can enhance the accuracy of short-term forecasting of air traffic flow.

Acknowledgements

This work was supported by the National Science Foundation of China (U1233101, 71271113, 60979018) and by the Fundamental Research Funds for the Central Universities(NS2016062).

References:

- [1] Faraway J, Chatfield C. Time series forecasting with neural networks: a comparative study using the airline data [J], *Applied Statistics*, 1998, 47(2): 231-250.
- [2] Ilona W, Heike R, Michael L. Wavelet -based Time Series Prediction for Air Traffic Data[C]. *Proceedings of SPIE-The International Society for Optical Engineering* 5266, *Wavelet Applications in Industrial Processing*, 2004: 238-248
- [3] CHENG Taoya, CUI Deguang, CHENG Peng. Data Mining for Air Traffic Flow forecasting: a Hybrid Model of Neural Network and Statistical Analysis[C]. *Intelligent Transportation Systems 2003 Proceedings*. 2003 IEEE: 211-215
- [4] CUI Deguang, WU Shuning, XU Bing. Air traffic flow forecasts based on artificial neural networks combined with regression methods[J], *Journal of Tsinghua University (Science and Technology)*, 2005, 45(1): 96-99
- [5] GENG Rui, CUI Deguang, XU Bin. Support vector machine-based combination model for air traffic forecasts[J], *Journal of Tsinghua University (Science and Technology)*, 2008, 48(7): 1205-1208
- [6] LIU Yu-mei. Air Traffic Flow Forecast Based on Least Square Estimate [J], *Journal of Civil Aviation University of China*, 2003, 21(4): 20-23
- [7] XU Xiao-hao, LIU Wei-xiang, WANG Xing-long, YUE Ren-tian. Study on Combined Prediction Method of Waypoint Flow Based on Genetic Algorithm[J]. *Journal of Civil Aviation University of China*, 2008, 26(6): 1-4
- [8] ZHAO Yi-fei, WANG Hong-yong. A Model for Short-term Prediction of Air Traffic Flow with Holt-Winters Model[J]. *Journal of Transportation Engineering and Information*, 2007, 5(4): 23-28
- [9] ZHAO Yu-huan, GUO Shuang. Study on Forecasting Model of Air Traffic Flow Considering Stochastic Factors[J], *Journal of Civil Aviation University of China*, 2008, 26(4): 59-61
- [10] Hansen L K, Salamon P. Neural Network Ensemble [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12 (10): 993-1001
- [11] Zhou Z-H, Wu J, Tang W. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*[J], *Artificial Intelligence*, 2002, 137 (1-2): 239-263
- [12] Zhang Zhaoning, Wang lili. *The theory and method of air traffic flow management*[M]. Beijing: Science Press, 2009: 32-55
- [13] Bermudez J D, Segura J V, Vercher E. Holt-Winters forecasting: an alternative formulation applied to UK air passenger data [J]. *Journal of Applied Statistics*, 2007, 34(9): 1075-1090
- [14] LU Junhui, Zhou Gang, in Yi. Adaptive aberrant network traffic detection algorithm based on time series forecast [J], *Journal of Beijing University of Aeronautics and Astronautics*, 2009, 35(5): 636-639