Air traffic complexity evaluation with novel complexity features and mRMR-XGBoost

Minghua Zhang¹*, Hua xie¹, Jiaming Ge², Dongfang Zhang²

¹ College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 210016, China
²College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 210016, China
*Corresponding author’s e-mail: mh.zhang@nuaa.edu.cn

Abstract. Air traffic complexity evaluation is a critical problem in air traffic system operation, especially for air traffic safety and air traffic controller deployment. Many researches focus on using mathematical modeling or machine learning methods to evaluate air traffic complexity. However, there are still challenges in accurate evaluation, which is affected by lack of effective complexity related features and complicated feature-complexity relationship. Based on the existing complexity features, this paper proposes additional domain features, time-dependent features and data distribution features to construct a novel air traffic complexity feature set. A mRMR method for feature selection is then applied to filter out redundant features. Finally, the filtered feature set and corresponding air traffic complexity level are input into XGBoost model for learning the relationship, so as to achieve high-performance evaluation of air traffic complexity in the face of new air traffic data. The experimental results show that the proposed features are beneficial to the evaluation of air traffic complexity, and the XGBoost model with mRMR method can effectively select the important features and mine the relationships within air traffic complexity data, resulting in an improvement in overall evaluation performance by at least 5% while using less than half of the original number of features.

1. Introduction

Nowadays, the air transport industry is developing rapidly with growing movement demand of goods and people, which results in the surging air traffic volume. For the safe and efficient management of the huge air traffic, the airspace is divided into smaller sectors to be managed separately, and each sector has a group of air traffic controllers (ATCos) responsible for directing the operation of aircrafts. However, the ATCos resource is limited and almost unable to meet the current increasingly busy air traffic. In order to keep the workload of ATCos within a safe range, we need to monitor the workload in real time and allocate ATCos resources over different sectors reasonably. To implement these, it is critical to accurately evaluate air traffic complexity.

The research on air traffic complexity evaluation is mainly divided into two major directions: mathematical modelling and machine learning. The former one characterizes air traffic complexity by constructing a mathematical model or the most relevant indicators[1], such as conflict probability, conflict resolution difficulty, Lyapunov Exponent and so on. The second approach, from the perspective of machine learning, regards air traffic complexity evaluation as a machine learning problem[2–4]. By considering different types of features that influence air traffic complexity as much
as possible and using machine learning model to learn the relationship between numerous features and traffic complexity, evaluation of air traffic complexity can be realized.

There are a large number of researches using various machine learning models to study air traffic complexity evaluation, but such researches mainly focus on the improvement of the model algorithm, and does not pay too much attention to complexity related features. On the one hand, the reason is that there are many factors influencing complexity characteristics, which are difficult to summarize and sort out; on the other hand, it is also because air traffic complexity is an abstract concept, and it is difficult to construct new effective features to describe air traffic complexity.

It is well known that the ceiling of machine learning model performance depends on the quantity and quality of input features. In order to further improve the performance of air traffic complexity evaluation, this paper summarizes and organizes the existing commonly used complexity features, while supplementing additional domain related complexity features. At the same time, it also proposes time-dependent features and data distribution features with the aim to comprehensively describe the information of air traffic complexity. To reduce the impact of redundant features, we introduce the mRMR method to screen the important features, and then uses an ensemble learning model (XGBoost) to mine the information between the screened features and the complexity labels to build a high-performance air traffic complexity evaluation model.

2. Air traffic complexity features analysis and novel features proposal

Air traffic complexity is abstract, which is difficult to be described and quantified in detail. In order to evaluate air traffic complexity, we need to describe and express it through relevant influencing factors. In this section, we will analyses the existing common complexity features, and then supplement other domain-related features. Meanwhile, time-dependent features and feature data distribution features will be proposed to provide a complete description of air traffic complexity.

2.1. Prevailing air traffic complexity features

Many scholars have explored and studied various features that can influence the level of air traffic complexity. At present, the commonly used air traffic complexity feature set is proposed by Gianazza and Gutted [5], which includes aircraft number information in different flight states or at different time periods (N, N², N₄₄, N₃₃, F₅, F₁₅, F₃₀, F₆₀, Dens), relevant parameters of aircraft movement rate (σₓ , σₓᵧ , avg_vs ), convergence and dispersion situation of aircraft groups (hpro₁ , vpro₁ , vpro₂ , int_hori , int_vert , track disorder , speed disorder ), difficulty degree of flight conflict resolution and conflict probability (Div, Conv, sensi_d, sensi_c, insen_d, insen_c, creed_ok, creed_pd ), etc. The above feature set consists of 28 features in total and has been consistently found to be important in reflecting air traffic complexity[2–4]. For a more thorough review of the above features, the cited source literatures can be referred.

2.2. Additional domain knowledge features

Apart from the complexity related factors considered in Section 2.1, there are other factors that can be taken into account in air traffic complexity evaluation. Since the change of aircraft motion state may cause the change of the entire air traffic situation, it is necessary to calculate the relevant parameters of the aircraft motion state change. Therefore, this paper adds the number of level flight aircraft, the number of aircraft whose heading/speed/altitude is in a state of change. In addition, the altitude, speed and corresponding control strategy of aircraft will be different with the different types of aircraft. In order to reflect the influence of the aircraft types, we use aircraft type mixing ratio as a complexity related feature. At the same time, the radiotelephony communication of the controller can reflect the workload of air traffic controllers, which is closely related to air traffic complexity, so the communication time and communication frequency characteristics are designed. In addition, we also add two features that could reflect vertically approaching trends according to the reference 6 [6].
2.3. **Time-dependent features**
Air traffic complexity is an embodiment of flight situation, which changes gradually and slowly in the huge airspace sector, so it has the characteristics of continuity, and air traffic complexity in adjacent periods has a certain correlation. Specifically, the complexity of the previous period may affect the complexity of the next period, and the two will have similar characteristics. Therefore, time-dependent features can be designed to capture the correlation of time series. In this paper, two kinds of time-related features are proposed, namely time information features (each complexity sample corresponds to the day of the week, the hour of the day, the minute of the hour) and the flight status in adjacent time periods (the number of aircraft in the heading/altitude/speed change state within 3 minutes before and after a certain sample, and the total time and frequency of radiotelephony communication within 3 minutes before and after).

2.4. **Data distribution features**
Although domain knowledge features and time-dependent features can provide actual business information, the amount of information may still have the potential to increase. In many practical machine learning tasks, exploratory analysis of data will be carried out based on the characteristics of actual data, which provides a basis for feature engineering. One of them is to extract new features according to the regular characteristics of data distribution. These features may have no practical meaning, but because of their certain data distribution rules and characteristics, they may play a certain role in improving the learning and training of the model. Therefore, this paper puts forward a benchmark distance feature based on the distribution of actual air traffic complexity feature data. Specifically, the first step is to draw a data distribution map between two features, and then distinguish different types of air traffic complexity samples with different colors, so as to observe the relationship between the correlation distribution of different features and the complexity category. The following two figures enumerate the feature data distribution relationship. It can be found that there is an obvious location difference between samples with high complexity and samples with lower complexity, which can be distinguished from spatial distribution. Therefore, this paper sets the midpoint of the most complex sample group as a benchmark, and the distance from the corresponding features of other samples to the benchmark is defined as the data distribution features. For our actual data, a total of 10 data distribution with these obvious characteristics are screened, and corresponding 10 data distribution features are constructed.

![Figure 1](image_url). The distribution of air traffic complexity feature data.

3. **XGBoost classification model with mRMR method**
A comprehensive feature set plays a great role in the task of machine learning, so we constructed a meticulous feature set from different perspectives in the previous section to describe air traffic complexity. But this does not mean that the more features the better. Redundant and useless features often occur and mislead the learning of the model, thereby affecting the predictive performance.
Therefore, on the one hand, we need to carefully review these features before input to the machine learning model to ensure the effectiveness; on the other hand, we need to use a robust and powerful model to mine the relationship within data to ensure excellence learning performance. So, in this section, we will introduce the mRMR feature selection method and XGBoost classification model used in this paper.

3.1. mRMR feature selection method

In view of information theory, the purpose of feature selection is to find a feature set \( S \) with \( m \) features \( F_i \), which jointly have the largest dependency on the target class \( c \). This scheme, called Max-Dependency, can be represented as \( I(\{F_i, \ i=1,2,...,m\}; c) \). But this criterion is hard to implement and time-consuming, an alternative is to select features based on maximal relevance criterion, which is called Max-Relevance. Max-Relevance is to search feature satisfying \( \max \frac{1}{|S|} \left( \sum_{F_i \in S} \text{Rel}(F_i; C) \right) \), which approximates Max-Dependency with the mean value of all mutual information values between every feature \( F_i \) and class \( c \). According to Max-Relevance, it is likely that features selected could have rich redundancy. When two features highly depend on each other, the respective class-discriminative power would not change much if one of them were removed. Therefore, Min-Redundancy criterion can be added to select mutually exclusive features, which is \( \min \frac{1}{|S|} \left( \sum_{F_i \neq F_j} \text{Red}(F_i; F_j) \right) \).

The criterion combining the above two constraints is called "minimal-redundancy-maximal-relevance (mRMR)"[7]. Combined with the above two criterions, the mRMR criterion can be expressed as \( \max \frac{1}{|S|} \left( \sum_{F_i \in S} \text{Rel}(F_i; C) \right) - \frac{1}{|S|^2} \sum_{F_i, F_j \in S} \text{Red}(F_i; F_j) \). In practice, incremental search methods can be used to find the near-optimal features set. Suppose we already have \( S_{m-1} \), the feature set with \( m-1 \) features. The task is to select the \( m \)-th feature from the set \( \{X - S_{m-1}\} \). The respective incremental algorithm optimizes the following condition:

\[
F_m = \arg \max_{F_i \in U} I(F_i; C) - \frac{1}{m-1} \sum_{F_i, F_j \in S_{m-1}} I(F_i; F_j) \quad (1)
\]

3.2. XGBoost classification model

XGBoost is short for “extreme gradient boosting”, which is designed to be a scalable machine learning system for tree boosting[8]. The parallel tree boosting and regularization strategy enable it to run in a much faster way and achieve state-of-the-art results in many machine learning problems. As an ensemble method, the basic idea of XGBoost is to combine several weak models into a strong one, which can be presented as:

\[
\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i) \quad (2)
\]

where \( f_k(\cdot) \) is a weak model and \( K \) is the number of weak models?

As a tree boost, the core of XGBoost is the Newton boosting, which searches the optimal parameters by driving the objective function as formula (8) towards the minimum direction.

\[
L(\phi) = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \frac{1}{2} \sum_{i=1}^{K} \Omega(f_i) \quad (3)
\]

\[
\Omega(f_i) = \gamma T + \frac{1}{2} a \| \phi \|_2^2 \quad (4)
\]
where $l$ is the loss function and $Q$ is the regularized term? They measure the performance and control the complexity of the model.

The ensemble model works better in an additive manner. $f_i$ is added to improve the model and the new objective function is formed as:

$$L' = \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) + Q(f_i)$$

(5)

where $\hat{y}_i^{(t)}$ is the prediction of the $i$-th sample and $f_i$ is the weaker model at the $t$-th iteration.

Several improvements are used in XGBoost to promote the classification performance, such as overfitting prevention, computation enhancement and so on, so we adopt it as our air traffic complexity learning model to mine information.

4. Experimental studies

4.1. Dataset preparation and evaluation metrics

In this section, to verify the effectiveness of our proposed air traffic complexity features and XGBoost classification model with mRMR method, several experiments are executed on the real air traffic operation data of a Guangzhou en-route airspace in China during December 1st to December 15th, 2019. After data pre-processing, checking and screening, we collected a total of 3605 valid samples, each of which corresponds to a one-minute air traffic scenario of the en-route airspace, consist of various complexity factors (50 features) as described in Section 2 and a corresponding complexity level (5 levels) obtained from air traffic management experts.

To evaluate the performance of different experiment, overall accuracy (Acc), mean absolute error (MAE) and F1-score are selected as the evaluation metrics. Their definition are as follows:

$$\text{accuracy}(\hat{y}, y) = \frac{\text{Counts} [\hat{y}_i = y_i]}{N}$$

(6)

$$\text{MAE}(\hat{y}, y) = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i|$$

(7)

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{Recall}}{\text{precision} + \text{Recall}}$$

(8)

4.2. Verifying the effectiveness of novel air traffic complexity features and XGBoost

As mentioned in Section 2, in addition to the prevailing complexity features (PF) used in existing research, this paper proposes different types of features, namely: additional domain features (ADF), time-dependent features (TDF), and data distribution features (DDF). In this part of the experiment, we carried out an ablation study, combining different types of features to study the effectiveness of our proposed features. According to the existing 4 types of features, we constructed a total of 8 complexity feature sets (FS_1: PF; FS_2: PF + ADF; FS_3: PF + TDF; FS_4: PF + DDF; FS_5: PF + ADF + TDF; FS_6: PF + ADF + DDF; FS_7: PF + TDF + DDF; FS_8: PF + ADF + TDF + DDF), in which FS is the abbreviation of feature set, and the number represents the serial number.

We divided the complete data set into training set and test set in a 4:1 ratio. In order to ensure the balance between samples of different categories, a stratified sampling method is adopted. K Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Gradient Boosting Decision Tree (GBDT) were used as comparative experiments.

The experimental results for evaluating the test set are shown in Table 1. Among each column in the table, the top two best results are highlighted in boldface, and the column of sum statistics the times that the corresponding features set obtains the top two best results, which reflects the merits and demerits of the feature set. Among each row, the best result is indicated with an underline.

---

5
Table 1. Air traffic complexity evaluation performance in different feature sets and classifiers.

| Feature Sets | Accuracy | KNN | LR  | DT  | RF  | SVM | MLP | GBDT | XGBoost | sum |
|--------------|----------|-----|-----|-----|-----|-----|-----|------|---------|-----|
| FS_1         | 72.09%   | 68.26% | 64.47% | 69.96% | 71.96% | 71.62% | 72.51% | 74.45% | 0       |
| FS_2         | 70.99%   | 68.93% | 64.47% | 69.40% | 72.82% | 71.98% | 72.98% | 75.70% | 0       |
| FS_3         | 76.29%   | 68.96% | 66.07% | 73.45% | 75.31% | 76.03% | 75.67% | 79.42% | 6       |
| FS_4         | 70.09%   | 68.99% | 63.44% | 69.54% | 71.51% | 71.79% | 73.34% | 74.26% | 0       |
| FS_5         | 75.01%   | 68.82% | 67.27% | 72.70% | 74.62% | 74.01% | 75.50% | 79.61% | 3       |
| FS_6         | 69.93%   | 69.68% | 63.49% | 71.18% | 72.09% | 71.51% | 72.79% | 75.31% | 1       |
| FS_7         | 75.40%   | 70.10% | 66.69% | 72.26% | 74.87% | 76.12% | 75.50% | 79.31% | 6       |
| FS_8         | 74.81%   | 69.60% | 65.97% | 71.82% | 74.54% | 74.23% | 74.98% | 79.94% | 1       |

| Feature Sets | MAE | KNN | LR  | DT  | RF  | SVM | MLP | GBDT | XGBoost | sum |
|--------------|-----|-----|-----|-----|-----|-----|-----|------|---------|-----|
| FS_1         | 0.2891 | 0.3212 | 0.3767 | 0.3082 | 0.2847 | 0.2896 | 0.2804 | 0.2641 | 0       |
| FS_2         | 0.2985 | 0.3154 | 0.3759 | 0.3146 | 0.2782 | 0.2907 | 0.2771 | 0.2510 | 0       |
| FS_3         | 0.2457 | 0.3151 | 0.3559 | 0.2758 | 0.2522 | 0.2483 | 0.2494 | 0.2119 | 5       |
| FS_4         | 0.3024 | 0.3151 | 0.3842 | 0.3126 | 0.2926 | 0.2918 | 0.2835 | 0.2652 | 0       |
| FS_5         | 0.2574 | 0.3162 | 0.3456 | 0.2805 | 0.2596 | 0.2693 | 0.2511 | 0.2094 | 4       |
| FS_6         | 0.3109 | 0.3087 | 0.3864 | 0.2962 | 0.2863 | 0.2952 | 0.2785 | 0.2541 | 1       |
| FS_7         | 0.2544 | 0.3040 | 0.3495 | 0.2835 | 0.2582 | 0.2483 | 0.2519 | 0.2139 | 5       |
| FS_8         | 0.2607 | 0.3096 | 0.3600 | 0.2887 | 0.2605 | 0.2702 | 0.2563 | 0.2064 | 1       |

We have following observations from the results above:

- From the perspective of the feature sets, that is, by observing the performance of different feature sets on the classifiers, we can find that some feature sets have poor performance in evaluating air traffic complexity. Specifically, FS_2/4, compared with other feature sets, does not get the top two results on any of the classifiers. However, some feature sets have better evaluation results, such as FS_3/5/7. According to the statistical data of the sum column, it can be seen that the value of FS_3 is the largest, which means that the top two results have been obtained by it on more than 6 classifiers. FS_5/7 follows, and achieves the top two results on 3-6 classifiers. Considering that FS_1 is the basic feature set, the poor performance of FS_4 reflects that the addition of data distribution features may not significantly improve the performance of the basic feature set; while FS_3/5/7/8 were found to contain time-dependent features, indicating that time-dependent information plays a vital role in evaluating air traffic complexity.

- From the perspective of the classifiers, that is, to observe the performance of different classifiers on the feature sets. It is found that XGBoost has achieved excellent results. No matter which feature set, it has achieved the best evaluation results compared with other classifiers. The optimal evaluation performance (Acc:79.94%, MAE:0.2064) is achieved under FS_8, which shows that XGBoost own strong feature adaptability and the learning ability. KNN, RF, SVM, LP, GBDT came to the second to the evaluation performance, while LR and DT were the worst.

- It is worth noting that FS_8 acquires the optimal results of all experiments on XGBoost but does not show excellent performance on other classifiers, in which FS_8 only achieves top two results on one classifier. This shows some features may be useless for some classifiers, which means that more features are not necessarily better. Key and important features have a great impact on the performance of the model. Redundant and inefficient features cannot or even are not conducive to the improvement of the evaluation performance, so it is necessary to carry out appropriate feature selection.

4.3. Performance study on XGBoost model with mRMR method

In order to improve the performance of the model, it is necessary to provide an effective feature set for the model. At the same time, feature selection can also greatly reduce the feature dimension, speed up model training and calculation, and can also provide assistance for the analysis of important features.
The mRMR method used in this paper is a feature selection method based on information theory, and shows excellent results on a large number of machine learning tasks.

To compare the effects of different feature selection methods, this paper selects another four feature selection methods (ReliefF, Backward, Gini index and CMIM) and random feature selection strategies, using XGBoost as the basic classifier to study their performance in evaluating air traffic complexity. The experimental results are shown in the Table 2, showing the optimal accuracy and corresponding number of features under different feature selection methods. In addition, we set a suboptimal threshold for different metrics to further observe the changes in the number of features.

| Method   | Acc(best)  | MAE(best)  | F1-score(best) | Acc(75%)  | MAE(0.25) | F1-score(70%) |
|----------|------------|------------|----------------|------------|-----------|----------------|
| Random   | 79.17% [n=46] | 0.2183 [n=46] | 70.89% [n=46] | 78.36% [n=41] | 0.2225 [n=42] | 70.89% [n=46] |
| ReliefF  | 79.39% [n=49] | 0.2122 [n=49] | 72.32% [n=49] | 78.58% [n=32] | 0.2200 [n=32] | 71.46% [n=30] |
| Backward | 79.50% [n=44] | 0.2119 [n=44] | 72.83% [n=23] | 78.00% [n=20] | 0.2266 [n=20] | 70.16% [n=20] |
| Gini     | 79.64% [n=49] | 0.2091 [n=49] | 73.93% [n=44] | 78.03% [n=35] | 0.2272 [n=45] | 70.07% [n=49] |
| CMIM     | 78.75% [n=50] | 0.2183 [n=50] | 71.72% [n=29] | 78.09% [n=43] | 0.2258 [n=44] | 70.00% [n=26] |
| mRMR     | 80.50% [n=21] | 0.1989 [n=21] | 71.10% [n=24] | 78.61% [n=19] | 0.2297 [n=17] | 70.50% [n=21] |

From the results in the table, it can be easily found that mRMR method has achieved the best results under all performance metrics, and is ahead of other feature selection methods, and the corresponding number of features is only 21. This means that only using less than half of the original number of features, we can achieve a better performance than the full feature set. For other feature selection methods, the use of feature selection methods does have some effects. While reducing the number of features used, they can still maintain the similar evaluation performance as the original full feature set. This shows that feature selection for air traffic complexity features is effective and has a prospect in practical engineering applications.

In order to further understand the intermediate process of the feature selection method, we provide the performance curves of different feature selection methods when the number of features changes. The left picture shows the change trend of Acc, the right picture shows the change trend of MAE, and the horizontal axis of the figure represents the number of features.

Figure 2. Evaluation performance curve of different feature selection methods.

As can be seen from Figure 2, in the overall trend, the evaluation performance of the model generally increases with the increase of the number of features, because adding more features is more likely to provide more information to the model. The mRMR method reaches the optimum when the number of selected features is 21, and then begins to decline slowly, indicating that the first 21 features provide extremely useful information, and the subsequent features are not enough to provide more information to help improve the performance of the learning model.

The Gini method is the fastest in terms of performance improvement, achieving accuracy of more than 75% in less than 10 features. The CMIM method is superior to the random method in the first 35-
40 feature numbers. Although the features selected in the later period continue to be beneficial to the model, they are inferior to the features selected by random method. The Backward method and ReliefF method are just the opposite. Before the number of feature selections is 17 and 27, the selected feature set is not as good as the random feature selection method, but it is reversed and maintained at a high level in the later stage.

5. Conclusion
In this paper, we propose domain knowledge features, time-dependent features and data distribution features that related with air traffic complexity. On the basis of the prevailing complexity features, the above new proposed features are synthesized to form a more complete air traffic complexity feature set used to characterize air traffic complexity. In order to ensure the effectiveness of the feature set, the mRMR method is introduced to perform feature selection to select the most useful features for machine learning model. Finally, we use XGBoost model to train and learn the data set after feature screening, and obtain the air traffic complexity evaluation model. The experimental results illustrate the effectiveness of our proposed complexity features and outstanding performance of XGBoost model with mRMR method.

In academic research, the proposed complexity features and feature selection methods in this paper can support scholars to analyses air traffic complexity at a deeper level, and help to excavate the internal mechanism of the formation of air traffic complexity. At the same time, in engineering application, the lower dimension complexity feature screened out by feature selection methods is helpful to be put into use and analysis of actual business, and can also reduce the calculation complexity to a certain extent.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (No.61903187) and Nanjing University of Aeronautics and Astronautics Graduate Innovation Base (Laboratory) Open Fund (kfjj20190732).

References
[1] Delahaye D, Paimblanc P, Puechmorel, Histon J.M, Hansman R.J. (2002) A New Air Traffic Complexity Metric Based on Dynamical System Modelization. In: The 21st Digital Avionics Systems Conference. Irvine. pp. 4A2:1-4A2:12.
[2] Zhu X, Cao X.B, Cai K.Q. (2017) Measuring air traffic complexity based on small samples. Chin. J. Aeronaut. 30:1493–1505.
[3] Cao X.B, Zhu X, Tian Z.C, Chen J, Wu D.P, Du W.B. (2018) A knowledge-transfer-based learning framework for airspace operation complexity evaluation. Transportation Research Part C: Emerging Technologies, 95:61–81.
[4] Xiao M.M, Zhang J, Cai K.Q, Cao X.B. (2016) ATCEM: a synthetic model for evaluating air traffic complexity. J. Adv. Transp, 50:315–325.
[5] Gianazza D, Guittet K. (2006) Selection and Evaluation of Air Traffic Complexity Metrics. In: IEEE/AIAA 25TH Digital Avionics Systems Conference. Portland. pp. 1–12.
[6] Chatterji G, Sridhar B. (2001) Measures for air traffic controller workload prediction. In: 1st AIAA, Aircraft, Technology Integration, and Operations Forum. Los Angeles. pp. 14.
[7] Peng H.C, Long F.H, Ding C. (2005) Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27:1226–1238.
[8] Chen T.Q, Guestrin C. (2016) XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco. pp. 785–794.