An Aircraft Trajectory Anomaly Detection Method Based on Deep Mixture Density Network

CHEN Lijing[1,2], ZENG Weili[1*], YANG Zhao[1]

1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P.R. China;
2. Aviation Key Laboratory of Science and Technology on Fault Diagnosis and Health Management, Shanghai 201601, P.R. China

Abstract: The timely and accurately detection of abnormal aircraft trajectory is critical to improving flight safety. However, the existing anomaly detection methods based on machine learning cannot well characterize the features of aircraft trajectories. Low anomaly detection accuracy still exists due to the high-dimensionality, heterogeneity and temporality of flight trajectory data. To this end, this paper proposes an abnormal trajectory detection method based on the deep mixture density network (DMDN) to detect flights with unusual data patterns and evaluate flight trajectory safety. The technique consists of two components: Utilization of the deep long short-term memory (LSTM) network to encode features of flight trajectories effectively, and parameterization of the statistical properties of flight trajectory using the Gaussian mixture model (GMM). Experiment results on Guangzhou Baiyun International Airport terminal airspace show that the proposed method can effectively capture the statistical patterns of aircraft trajectories. The model can detect abnormal flights with elevated risks and its performance is superior to two mainstream methods. The proposed model can be used as an assistant decision-making tool for air traffic controllers.

Key words: aircraft trajectory; anomaly detection; mixture density network; long short-term memory (LSTM); Gaussian mixture model (GMM)

0 Introduction

The contradiction between limited airspace resources and the rapidly increasing demand for air traffic has become increasingly prominent, leading to increased flight safety issues, especially in some congested airspace. Timely detection of abnormal flight trajectory is vital to prompt response of air traffic controller (ATC), and safety risk. According to statistics released by Boeing[1], 78% of fatal air-related accidents worldwide occurred in the terminal airspace, including departure and arrival flight phases of the aircraft. Therefore, developing new anomaly detection technologies matters to the analysis of abnormal behavior of aircraft in terminal airspace[2]. The main aim of this paper is to construct a trajectory anomaly detection method in terminal airspace.

Regarding the trajectory anomaly detection, many methods based on movement trajectories have been proposed in the aviation domain[3-7]. In recent years, researchers can access vast amounts of data via open or public sources like ADS-B[8]. The advances in machine learning have spurred the research's pace to detect unsafe flight patterns and other operational anomalies automatically.

Clustering-based approaches have been extensively utilized in anomaly detection. Barratt et al.[9] used K-means to cluster the trajectories and construct a probabilistic model based on a Gaussian mixture model. The clusters with low probability were regarded as anomalous trajectories. Guo et

al.[10] presented an unsupervised clustering algorithm based on the information bottleneck (IB) principle and mutual information (MI). The kernel density estimation was used to model the shape of the motion trajectory. The Shannon entropy was adopted to identify whether a test trajectory was anomaly or not. Gariel et al.[11] identified each trajectory's turning point and clustered all turning points using density-based spatial clustering of applications with noise (DBSCAN), and then identified anomalous trajectories. These algorithms need to measure trajectory distances[12], such as Euclidian distance, the dynamic time warping and the Hausdorff distance, and cost more money, time and high-performance equipment when applied to large-scale datasets. Actually, it is challenging to obtain a good measure of the trajectory distance. Besides, clustering-based approaches do not scale well with vast amounts of high-dimensional data. Therefore, they are not suitable for real-time anomaly detection.

Domain-based method is one category of the state-of-the-art methods for detecting significant events in aviation. Das et al.[13] applied the multiple kernel anomaly detection to detect the approach phase. They correctly identified two anomalous situations, that is, high energy approaches and turbulent approaches. Puranik et al.[14] proposed a framework to identify abnormal trajectories based on one class support vector machine (OC-SVM) model that can compute the anomaly scores of each trajectory. The method was assessed by simulated flight data with abnormalities and real flight data during the departure and arrival phases. The model showed the excellent result in anomalous detection, while its computational complexity was expensive and unsuitable for large-scale datasets. To reduce the time of learning on large datasets, the extreme learning machines (ELM)[15] or the local outlier probability (LOP) method[16] will be a better choice.

Some advanced methods are proposed in anomaly detection because of the development of neural networks. In principle, deep-learning methods should be better than traditional machine learning methods[17], especially in the case of finding anomalies in large-scale data. To better identify the operation mode, many scholars used neural networks to learn trajectory characteristics. These methods assume that anomalies will lose information when projected to a lower dimension space, resulting in a relatively poor reconstruction result. Autoencoders (AE) has the excellent capability of nonlinear dimension reduction[18]. Olive et al.[19] combined a trajectory-clustering method to obtain the main flows in airspace with autoencoding neural networks to perform anomaly detection in flown trajectories. The author used a clustering method to recognize a group of clusters corresponding to sector flows. Then AE was used in each cluster to detect abnormal trajectories. Aircraft trajectory related information belonged to the time series data with high spatial-temporal correlation. The recurrent neural network (RNN) was used to capture temporal and nonlinear dependencies in multivariate time series as the one available in the aircraft trajectory dataset[20]. The long short-term memory is a better tool than RNN because it can solve the inability to learn long term patterns in sequential data[21]. Malhotra et al.[22] proposed a long short-term memory (LSTM) encoder-decoder model, that was trained only with normal data. The trajectories with high reconstruction errors were classified as abnormal trajectories. Nevertheless, the network output only a limited description of the target variables' properties, instead of sufficient analysis of abnormal trajectories. All of these methods are for purpose of learning features from a huge scale of data. Optimizing the trajectory features learning ability and the models' outputs to describe target variables is the main direction.

The key limitations of the studies motioned above are: (1) Their computational costs are high and it is difficult to apply them to large-scale data sets; (2) aircraft flight data has high spatial-temporal correlation, which is challenging for learning the attributive characteristics; (3) the limited description of the target variables' properties does not pro-

vide sufficient analysis on abnormal trajectories. To address these problems, this paper builds a novel detection model based on a deep mixture density network (DMDN) with the following contributions: (1) A regularization model is utilized to handle the difficulties in trajectory reconstruction and obtain a smooth trajectory with less noises; (2) we train a deep LSTM to capture features of a large number of trajectories and take into account the attributes closely related to aircraft trajectory status, such as aircraft type, speed, and angle, thereby obtain a richer representation of trajectories; (3) the model uses the Gaussian mixture model (GMM) to parameterize the statistical properties of the flight trajecto-

ry. The output of the model can characterize the trajectory data, so as to evaluate the risk degree of trajectory.

1　Methodology

Our model first uses a regularization method to get a high quality of trajectories sequences by time-aligning trajectories and extrapolating the short trajectories. The DMDN is then introduced to characterize trajectories. Finally, the conditional probability density function of the trajectories can be used to assess the risk of trajectory. Fig.1 shows the workflow of the trajectory anomaly detection method.
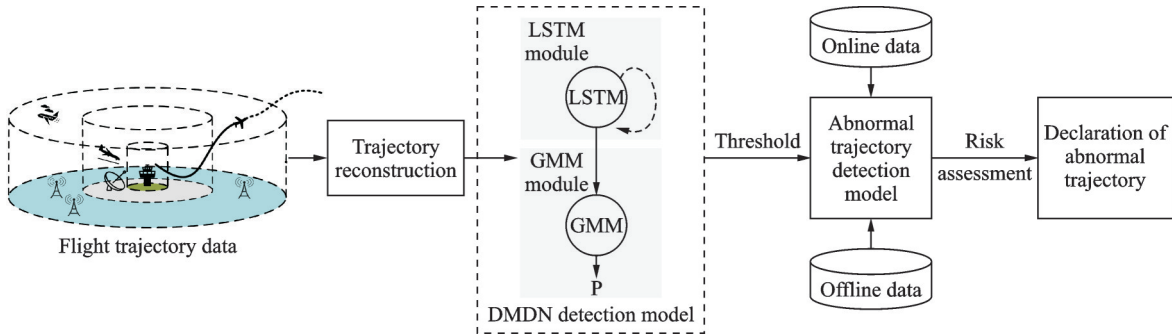

Fig.1　Flowchart of the trajectory anomaly detection method

1. 1　Trajectory reconstruction

By quality analysis of the flight data, it is found that there exist mainly two problems of the original data. First, some trajectories may have outliers (such as noises, null values) at a given time step. Second, the lengths of time-gaps within a trajectory differs from point to point. The regularization method is a widely used framework in image reconstruction, denoising , data compression, etc. Trajectory reconstruction employs the regularization method and uses a low-quality trajectory sequence to produce a higher quality trajectory. Here, we tackle the two problems existing in the original trajectories in a unified model.

Define T_{com} with the unit of second to present the common length. The method extrapolates trajectories that are shorter than T_{com} and truncates trajectories that are longer than T_{com}. An optimal value of T_{com} can be obtained through experiments from historical flight data.

The regularization method for aircraft trajectory reconstruction can be described as the following optimization problem

$$\bar{P} = \operatorname*{ArgMin}_{P}\left\{\left\|DP - P_0\right\|_{\mathrm{F}}^2 + \lambda\rho(P)\right\} \qquad (1)$$

where $\bar{P}$ is the reconstructed trajectories; $P \in R^{N \times E}$ the high-quality trajectory with N equal time-gap points and its ith row is the trajectory point with attributes of E dimensions at time i; length N the maximum of the trajectory original length and T_{com}. $P_0 \in R^{H \times E}$ the corresponding observed trajectory with H unequal time-gap points; $D \in R^{N \times N}$ a sampling matrix; $\left\|DP - P_0\right\|_{\mathrm{F}}^2$ represents a measure of fit; $\rho(\cdot)$ the regularization cost function, used to adjust the smoothness of the trajectory; λ the regularization parameter used to control the trade-off between the fit term and a regularization term, and $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm of a matrix. Eq.(1) propels to smooth the measured points and is especially useful for suppressing noises.

The most useful property of a regularization function should have a good balance between removing noises and preserving the trajectory's local changing tendency. There are many formulations for the regularization term[23]. For the sake of simplicity, this paper adopts the Tikhonov cost function, one of the most widely referenced regularization cost function in single processing, as the regularization term, and is defined as

$$\rho(\boldsymbol{P}) = \left\| \Delta \boldsymbol{P} \right\|_F^2 \qquad (2)$$

If the discrete form of Laplacian operator is approximated by

$$(\boldsymbol{\Gamma P})_i = \boldsymbol{P}_{i-1} - 2\boldsymbol{P}_i + \boldsymbol{P}_{i+1} \qquad (3)$$

where $\boldsymbol{\Gamma}$ is the Laplacian matrix. Then the Laplacian matrix looks like

$$\boldsymbol{\Gamma} = \begin{bmatrix} -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -2 \end{bmatrix}_{N \times N} \qquad (4)$$

The intuition behind this Tikhonov regulation term is to force the local spatial smoothness of the trajectory sequence. As the noisy points contain high-frequency energy, they will be removed in the regularization process, and the reconstructed trajectory will not include outliers.

Substituting the regularization term in Eq.(2) into Eq.(3), the trajectory reconstruction problem can be transformed into minimizing the following objective function

$$f(\boldsymbol{P}) = \left\| \boldsymbol{DP} - \boldsymbol{P}_0 \right\|_F^2 + \lambda \left\| \boldsymbol{\Gamma P} \right\|_F^2 \qquad (5)$$

The objective function Eq.(5) is a convex quadratic function and can be solved analytically. A solution $\boldsymbol{P}$ minimizes f if and only if

$$\nabla f(\boldsymbol{P}) = 0 \qquad (6)$$

The solution of convex quadratic problem is

$$\boldsymbol{P} = (\boldsymbol{D}^T \boldsymbol{D} + \lambda \boldsymbol{\Gamma}^T \boldsymbol{\Gamma}_0)^{-1} \boldsymbol{D}^T \boldsymbol{P}_0 \qquad (7)$$

The reconstructed trajectory data has the following properties：(1) The time intervals between any two continuous points are same；(2) the time resolution is higher than the raw trajectory；(3) the reconstructed trajectory is smoother than the original one. Fig.2 shows the results of trajectory reconstruction.
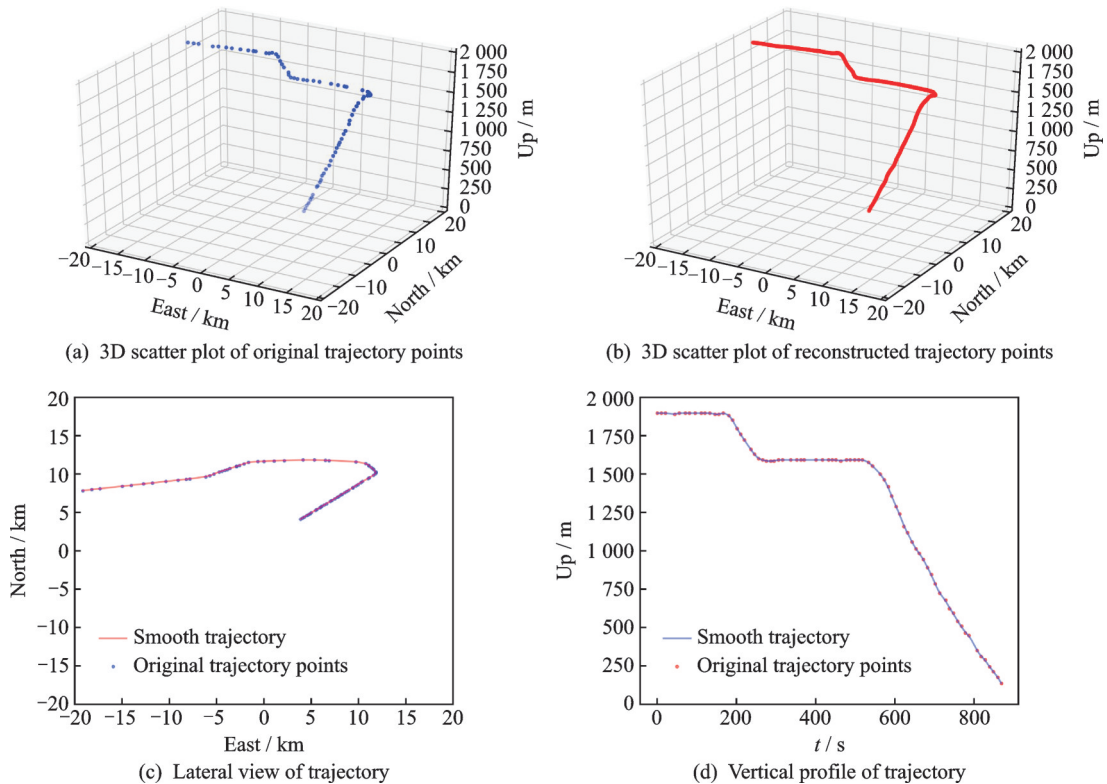

(a) 3D scatter plot of original trajectory points

(b) 3D scatter plot of reconstructed trajectory points

(c) Lateral view of trajectory

(d) Vertical profile of trajectory

Fig.2   Trajectory reconstruction effect

1. 2   Deep mixture density network

The mixture density network (MDN) proposed by Bishop[24] is a neural network that maps an input feature to a set of mixture density model parameters. MDN performs well in speech synthesis and written synthesis because it can well fit the time-varying features. Unlike conventional networks, MDN does not directly output a point estimation for the target data, but parameters of a multi-model mixture of distributions. It can better express the trajectory's statistical properties.

MDN is a combination of a neural network and a mixture density model, where the output of a neural network is the input parameters of a density mixture model. If we learn the trajectory features through the neural network with strong nonlinear mapping capability and then parameterize the data through the mixture density model, we can fit the trajectory data statistics well to determine whether the flight trajectory has abnormal characteristics.

In previous studies, shallow neural networks, like back propagation (BP) network, is utilized to learn features, but it could not fit complex time series data. Since aircraft trajectory information is time series data with high spatial-temporal correlation, this paper chooses LSTM network to learn the multidimensional coupled nonlinear characteristics of aircraft trajectories. It proves to be useful for learning feature related to time-sequential data. There are various probability density distributions, and GMM is chosen to parameterize the network because of the Gaussian distribution's good computational properties.

The output of the LSTM network determines the parameters of GMM. Therefore, DMDN represents the conditional probability density function of the trajectories conditioned on the input vector of the LSTM network. DMDN can model the underlying generator of flight trajectory data and identify latent risks from daily operations. Thus, we can use the results to judge abnormal trajectory.

（1）LSTM network

LSTM is a variant of RNN and composed of a cell and three gates, namely, an input gate, a forget gate, and an output gate. LSTM can maintain a longer memory, because it can decide which information to retain and which to forget. LSTM can consider the relationship between the input information in the past moment and the current moment, so it is suitable for modeling long time series data. Fig. 3 shows that the LSTM network with J hidden layers is trained to capture the characteristics of the input data. The updating process of the internal parameters in the LSTM network is mainly realized through the information transfer of the internal cell state and the hidden layer state in time step. The calculation formula is as follows

$$\begin{cases} f_t = \sigma(W_{xf}x_t + W_{sf}S_{t-1} + b_f) \\ i_t = \sigma(W_{xi}x_t + W_{si}S_{t-1} + b_i) \\ o_t = \sigma(W_{xo}x_t + W_{so}S_{t-1} + b_o) \\ C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{sc}S_{t-1} + W_{xc}x_t + b_c) \\ S_t = o_t \odot \tanh(C_t) \end{cases}$$

$$(8)$$

where x_t is the input data at time t; $\sigma(\cdot)$ the Sigmoid activation function; $\tanh(\cdot)$ a hyperbolic tan-
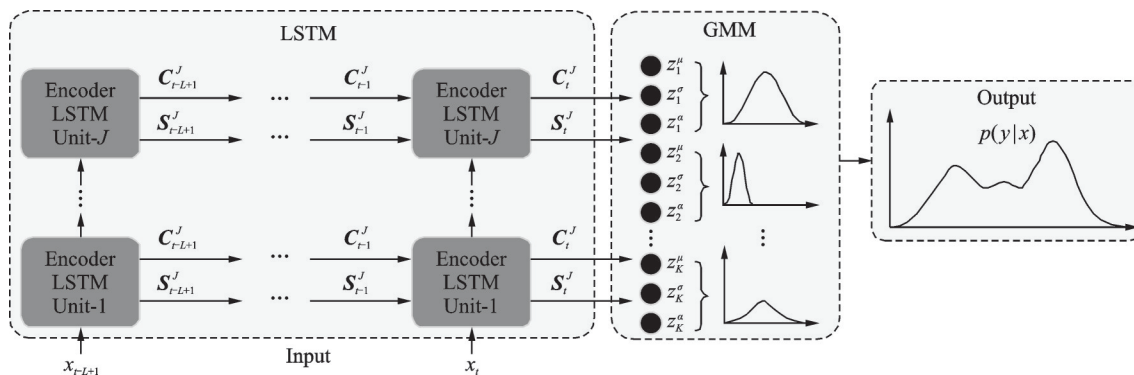

Fig.3   Deep mixture density network model

gent activation function; $\odot$ the Hadamard product of the matrix; f_t the control unit of the "forgetting gate"; i_t the control unit of the "input gate"; o_t the control unit of the "output gate"; C_t the cell state at time t; S_t the hidden state at time t; and W_{xf}, W_{sf}, b_f, W_{xi}, W_{si}, b_i, W_{xo}, W_{so}, b_o, W_{sc}, W_{xc}, b_c are the weight parameters in the network.

(2) GMM

GMM can be used for clustering and density estimation. We can identify frequent flight operations and calculate the statistical inference of the underlying distributions, representing the degree of abnormality of flight trajectories. The outputs of GMM are parametric probability functions, and represented as a weighted sum of Gaussian component densities. The outputs of LSTM network are utilized as the input of GMM. Let $\boldsymbol{Z}$ be the output tensor of MDN; It is partitioned into three subsets z_i^μ, z_i^σ and z_i^a that correspond to the inputs of the GMM mean, variances and weights, respectively

$$\boldsymbol{Z} = [z_1^\mu, \cdots, z_K^\mu, z_1^\sigma, \cdots, z_K^\sigma, z_1^a, \cdots, z_K^a] \qquad (9)$$

where K is the number of Gaussians. These parameters have some restrictions which should be satisfied

(1) The mixing coefficient α_i should satisfy $\sum_{i=1}^{K} \alpha_i = 1$. The mixing coefficients are estimated using the softmax activation function

$$\alpha_i = \frac{\exp(z_i^a)}{\sum\limits_{i=1}^{K} \exp(z_i^a)} \qquad (10)$$

where the value of α_i lies in the range $(0,1)$.

(2) The scale parameters σ_i should satisfy the constrain $\sigma_i(x) > 0$. It is defined as

$$\sigma_i(x) = \exp(z_i^\sigma) \qquad (11)$$

(3) The mean parameters μ_i represents location parameters. These should be directly dependent on the network outputs. It is defined as

$$\mu_i = z_i^\mu \qquad (12)$$

The output of the MDN model is the conditional density function of the target variable versus the input variable in the form

$$p(y|x) = \sum_{i=1}^{M} \alpha_i(x) N_i(y|x) \qquad (13)$$

where x is the input variable; y the target variable; M the number of components in the model; $\alpha_i(x)$ the mixing coefficient; and $N_i(\cdot)$ the kernel function.

The Gaussian function is chosen as our kernel functions

$$N_i(y|x) = \frac{1}{(2\pi)^{c/2} \sigma_i^c(x)} \exp\left\{ -\frac{\|y - \mu_i(x)\|^2}{2\sigma_i^2(x)} \right\} \qquad (14)$$

where $\mu_i(x)$ and $\sigma_i^2(x)$ are the mean and the variances of the ith Gaussian, respectively; and c is the dimension of the target variable to be obtained. We assume that each component of the output vector is independent of each other.

To make the output as close to the target variable as possible, we define a loss function that requires minimization of the negative logarithm of the likelihood

$$E = -\log(p(y|x)) = -\log\left(\sum_{i=1}^{M} \alpha_i(x) N_i(y|x) \right) \qquad (15)$$

The parameters in MDN will be optimized as the negative log-likelihood in Eq. (15) minimized. We choose the Adam optimizer function, an extension of the gradient descent method, to improve the traditional gradient by introducing the first-moment estimation and second-moment estimation. If we choose the mixing coefficients and the Gaussian parameters (means and variances) correctly, the MDN can approximate any given density function to arbitrary accuracy. Thus, we estimate the distribution of aircraft trajectory using the optimized MDN.

1.3 Anomaly detection

"Anomaly" means unusual. That means different flight operations with normal trajectories. The probability of its occurrence is much lower than normal. The detection of the anomalous trajectory is of interest to the aviation domain, and can be used as valuable analysis reference for airspace safety. For instance, some flight operations that may lead to safety incidents, such as go-around operations, conflict resolution activities, and air traffic management operations, require increased safety precautions. Fig.4 shows that the abnormal trajectories have dif-

ferent flight operations with normal trajectories, leading to more uncertainties and increased safety risks. This paper aims to determine a risk threshold that distinguishes abnormal trajectories from normal ones. Since the boundary between normal and abnormal is vague, we adopt a data-driven approach to find the threshold.
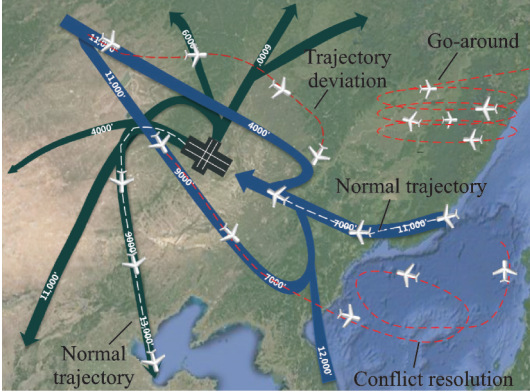

Fig.4　Schematic diagram of normal and abnormal trajectories

The paper obtains the threshold by training the normal flight dataset. Abnormal trajectories are detected if their probabilities of being normal are lower than the threshold. Mathematically, it is a sum of the probability density function of a trajectory belonging to a Gaussian component and weighted by the probability density function of the Gaussian component being appropriate in the form as

$$P(\boldsymbol{x}^f \text{ is normal}) = \sum_{i=1}^{M} p(i) \, p(\boldsymbol{x}^f \text{ is from } i) \quad (16)$$

where $\boldsymbol{x}^f$ is a vector form, trajectory f, and i one of the gaussian components

$$p(i) = \alpha_i \quad (17)$$

$$p(\boldsymbol{x}^f \text{ is from } i) = N(\boldsymbol{x}^f | \mu_i, \sigma_i) \quad (18)$$

In practice, air traffic controllers (ATCs) can adjust the value of threshold according to their preferences and needs. This value is associated with an air traffic detection system's sensitivity and can be set based on the distribution of all trajectories in the dataset. Therefore, the data-driven model is suitable for anomaly detection in different terminal areas.

If the probability of one trajectory is lower than the threshold, an ATC checks additional information about the trajectory, such as procedure standards and weather information, to further determine its risk. The detection threshold can be adjusted according to the actual situation. Therefore, this model can play an auxiliary decision-making role for air traffic control well.

2　Case Study

Our method is evaluated on departure and arrival flight datasets at Guangzhou Baiyun International Airport (GBIA).

2.1　Data description

The dataset used in this paper contains aircraft monitoring data of flights from GBIA in China, from November 2018 to December 2018, including 60 000 trajectories with 32 000 departure trajectories and 28 000 arrival ones. Each item of data consists of ten attributes, including flight ID, latitude, longitude, altitude, airspeed, course, departure location, arrival location, aircraft type, and monitoring time. The center of runways in GBIA is set to the earth-centered, earth-fixed coordinates. The original measurements are transformed into east-north-up coordinates. We only keep the position measurements that are less than 50 km in the "east" and "north" dimensions and less than 2 000 m in the "up" dimension, which is the critical area for tackoff and final approach.

Considering the anomaly detection model's training, this study transforms the raw flight data into vectors in the form of multivariate time-series. Each trajectory is represented by a series of time, position measurements, airspeed, course, and flight information (departure location, arrival location, aircraft type) ordered data. Among them each trajectory $\boldsymbol{\tau} = \{(t_i, \hat{\boldsymbol{p}}_i, v, \theta, \boldsymbol{D})\}_{i=1}^{N}$ where time $t_i \in \mathbf{Z}^+$, the measurements $\hat{\boldsymbol{p}}_i \in \mathbf{R}^3$, the airspeed $v \in \mathbf{R}$, the course $\theta \in \mathbf{R}$, and the flight information $\boldsymbol{D}$ is three-dimensional discrete random variables. The dataset is divided into a training set, the validation set and the test set according to the ratio of $8:1:1$. The test dataset with 56 abnormal departure trajectories and 72 abnormal arrival trajectories is used to check the model's quality.

The MDN model is run on a high-performance computer with 16 i7 CPUs and 128 GB RAM.

2. 2   Parameter selection

We carry out the experiment on the validation dataset with 66 abnormal trajectories in the departure validation dataset and 53 abnormal trajectories in the arrival validation dataset. The abnormal rate for the dataset is about 2%, meaning that the two classes (normal and abnormal) are disproportionally represented as a typical unbalanced dataset in the binary classification problem. Thus, the area under the curve (AUC) is a better index for assessing the model's performance, since it is not sensitive to sample categories' equilibrium.

The value of T_{com} has considerable influence on the model. For instance, a large T_{com} results in less accurate trajectories, as more trajectories are extrapolated and more trajectories are too short. Fig. 5 shows the histogram of the original arrival trajectory. We select T_{com} as the median of the trajectory lengths at GBIA: T_{com} is set to 190 s for departures and 550 s for arrivals.
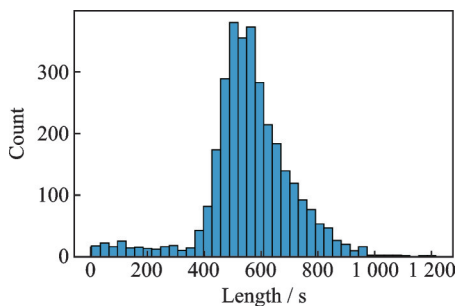

Fig.5   Histogram of raw arrival trajectory

The DMDN model has an input layer, the LSTM layer, the GMM layer, and an output layer. The input data contains seven-dimensional features. The LSTM layer with three hidden layers and the node of each layer is set to 128. Fig.6 shows the sensitivity analysis of the number of mixture components M based on AUC. The number of mixture components M is set to 15 for departures and 20 for arrivals. The output of MDN is set to three times of M. The model is trained using an Adam optimizer with an initial learning rate of 0.006.
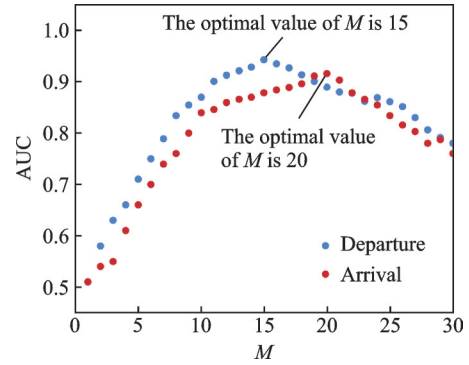

Fig.6   Sensitivity analysis of the number of mixture components based on AUC

The threshold of anomaly detection is associated with the sensitivity of the detection model. We used the receiver operating characteristic (ROC) curve to obtain the optimal threshold, since it is easy to detect the effect of arbitrary thresholds on the classifier's generalization performance and not affected by the distribution of the positive and negative samples. Fig.7 presents the ROC curve of our model in the arrival validation dataset. On the ROC curve, the point closest to the top of the coordinate diagram is the best threshold with high true positive rate and false positive rate.
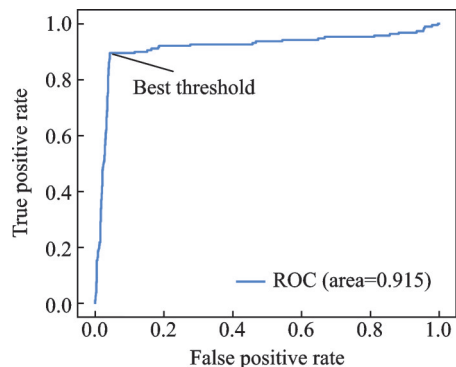

Fig.7   ROC curve of the proposed model in the arrival validation dataset

2. 3   Performance measure

To assess the performance of the detection method, we use four classical criteria, including precision, recall, F_{score}, and AUC. The definition are as follows

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (19)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (20)$$

$$F_{\text{score}} = (1+\beta^2)\frac{\text{Precision}\cdot\text{Recall}}{\beta^2\cdot\text{Precision} + \text{Recall}} \quad (21)$$

"Precision" represents the probability that a detected anomaly is indeed an anomaly. "Recall" represents the percentage of all real anomalies that are detected. Since the aircraft behavior could endanger human safety, all anomalies should be detected as far as possible. Therefore, recall matters more than precision in our task, so β is set to 2, which means recall has more weight than precision. The AUC concept is described in Section 2.2. The closer the AUC is to 1, the better the performance of model.

2.4　Experimental results

Fig.8 shows the results of 20 flight modes of arrival trajectories. The arrival aircraft follow the instrument landing system (ILS), which proves the validity of our data. It can be intuitively seen in Fig.9 that abnormal trajectories differ markedly from most normal trajectories. Some of the anomal trajectories do not deviate much from the normal ones, while in the lateral profile, the trajectories deviate significantly from the other trajectories in the same operation.

The proposed DMDN detection method is compared with a simple mixture MDN employing the multi-layer perceptron as the neural network structure[9]. Fig.10 shows the experimental results. Our method has the optimal in four assessment criteria, not only in departure trajectories but also in arrival trajectories. The difference between DMDN and MDN lies in the structure of the neural network layer. The performance of DMDN is better than that of MDN because of the excellent trajectory feature coding ability of LSTM. At the same time, learning probabilistic (LP) uses a classical clustering algorithm, like K-means, and is limited in its ability to characterize input characteristics.

By comparing Fig.10(a) and Fig.10(b), we can find that the departure trajectory detection model performs better than the arrival's trajectory anomaly detection model. Since a departure flight typically climbs out of the airport terminal area following a prescribed procedure, and its flight patterns are relatively simple and easier to mine from the data. However, an arrival flight will stay in the airport terminal area for a longer time, and its flight trajectories can change more randomly. Arrival flights may have
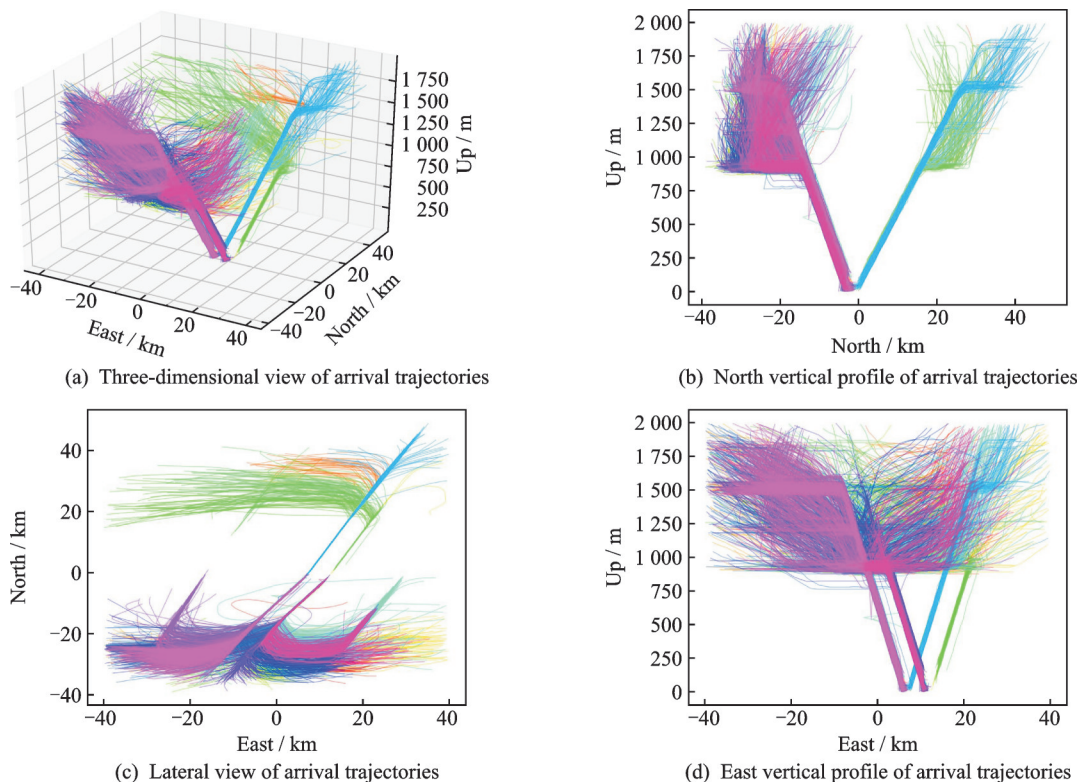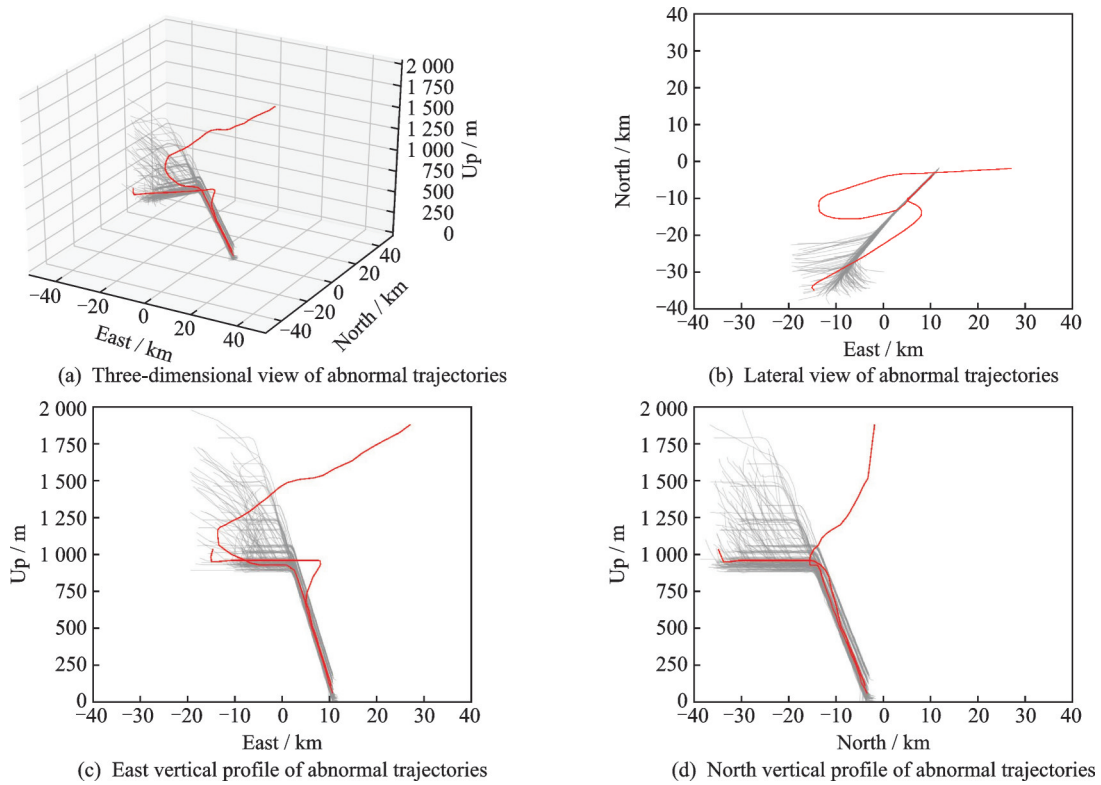

(a) Three-dimensional view of arrival trajectories

(b) North vertical profile of arrival trajectories

(c) Lateral view of arrival trajectories

(d) East vertical profile of arrival trajectories

Fig.8　Flight modes of arrival trajectories

(a) Three-dimensional view of abnormal trajectories

(b) Lateral view of abnormal trajectories

(c) East vertical profile of abnormal trajectories

(d) North vertical profile of abnormal trajectories

Fig.9　Visualization of abnormal arrival trajectories in a flight mode
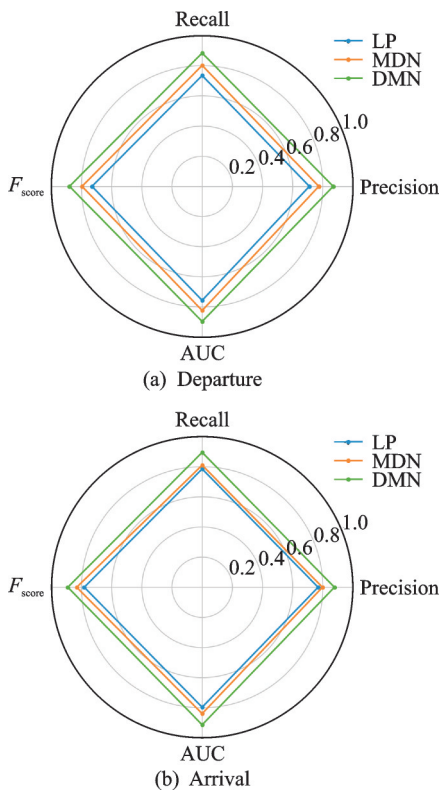

(a) Departure

(b) Arrival

Fig.10　Anomaly detection results of the three methods

air traffic control measures such as queuing for approach. So it is more challenging for the model to capture its trajectory characteristics.

In a nutshell, our trajectory anomaly detection method works best and can well capture the statistical features of aircraft trajectories. Also, our method trains quickly, compact, and is suitable for large-scale datasets.

3　Conclusions

This paper aims to build a data-driven method to improve the abnormal behavior detection system of aircraft to guarantee flight safety. Different from other data-driven methods, the proposed method considers not only the general position information and velocity of aircraft but also the multidimensional features such as angle and aircraft type to indirectly represent the flight intention and performance of aircraft trajectory. This is helpful to improve the effectiveness of the model for the extraction of the trajectory characteristics. Before this study, it was difficult to reconstruct the trajectory accurately based on the flight data, while our model can handle the difficulties in trajectory reconstruction and obtain a smooth trajectory with less noise. Our model is also useful to other vehicle's trajectory anomaly detection. It can be trained quickly, compact, and is suitable for large-scale datasets.

Several improvements to the proposed method will be performed in our future work. More aircraft attributes will be considered, such as the special position, ATC factors. Also, different terminal areas are different, so we should convert the threshold's training method to change the threshold in real-time. We will use a powerful visual analytics technique to visualize anomaly detection system interface and strengthen all the steps of the trajectories analysis.

References

[1] Boeing. Statistical summary of commercial jet airplane accidents[R]. USA: Boeing, 2015:23-25.

[2] LAZAREVIC A, ERTOZ L, KUMAR V. A comparative study of anomaly detection schemes in network intrusion detection[C]//Proceedings of the SIAM International Conference on Data Mining.[S.l.]: Society for Industrial and Applied Mathematics, 2003: 25-42.

[3] XU X, LIU H, YAO M. Recent progress of anomaly detection[J]. Complexity, 2019, 2019:1-11.

[4] AKOGLU L, TONG H, KOUTRA D. Graph based anomaly detection and description: A survey[J]. Data Mining and Knowledge Discovery, 2015, 29(3): 626-688.

[5] KUMAR V C A B. Anomaly detection: A survey[J]. ACM Computer Surveys, 2009,41:1-58.

[6] MARKOU M, SINGH S. Novelty detection: A review—Part 1: Statistical approaches[J]. Signal Processing, 2003,83(12):2481-2497.

[7] MARKOU M, SINGH S. Novelty detection: A review—Part 2[J]. Signal Processing, 2003, 83(12): 2499-2521.

[8] GAVRILOVSKI A, JIMENEZ H, MAVRIS D N, et al. Challenges and opportunities in flight data mining: A review of the state of the art[C]//Proceedings of AIAA SciTech 2016. San Diego, USA: AIAA, 2016:25-36.

[9] BARRATT S T, KOCHENDERFER M J, BOYD S P. Learning probabilistic trajectory models of aircraft in terminal airspace from position data[J]. IEEE Transactions on Intelligent Transportation Systems, 2018,20(9):3536-3545.

[10] GUO Y, XU Q, LI P, et al. Trajectory shape analysis and anomaly detection utilizing information theory tools[J]. Entropy, 2017,19(7):323.

[11] GARIEL M, SRIVASTAVA A N, FERON E. Trajectory clustering and an application to airspace monitoring[J]. IEEE Transactions on Intelligent Transportation Systems, 2011,12(4):1511-1524.

[12] ATEV S, MILLER G, PAPANIKOLOPOULOS N P. Clustering of vehicle trajectories[J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 3 (11):647-657.

[13] DAS S, MATTHEWS B L, LAWRENCE R. Fleet level anomaly detection of aviation safety data[C]// Proceedings of 2011 IEEE Conference on Prognostics and Health Management (PHM). Denver, USA: IEEE, 2011:1-10.

[14] PURANIK T G, MAVRIS D N. Anomaly detection in general-aviation operations using energy metrics and flight-data records[J]. Journal of Aerospace Information Systems, 2018,15(1):22-35.

[15] JANAKIRAMAN V M, NIELSEN D. Anomaly detection in aviation data using extreme learning machines[C]//Proceedings of International Joint Conference on Neural Networks. Vancouver, Canada: IEEE, 2016:1023-1042.

[16] OEHLING J, BARRY D J. Using machine learning methods in airline flight data monitoring to generate new operational safety knowledge from existing data[J]. Safety Science, 2019,114:89-104.

[17] CHAWLA R, CHAWLA S. Deep learning for anomaly detection: A survey[EB/OL]. (2019-01-23) [2020-10-03]. https://arxiv.org/abs/1901.03407.

[18] PIMENTEL M A F, CLIFTON D A, CLIFTON L, et al. A review of novelty detection[J]. Signal Processing, 2014,99:215-249.

[19] OLIVE X, GRIGNARD J, DUBOT T, et al. Detecting controllers' actions in past mode S data by autoencoder-based anomaly detection[C]//Proceedings of the 8th SESAR Innovation Days. [S.l.]: [s.n.], 2018:1-8.

[20] NANDURI A, SHERRY L. Anomaly detection in aircraft data using recurrent neural networks (RNN)[C]//Proceedings of 2016 Integrated Communications Navigation & Surveillance. Herndon, USA: IEEE, 2016:1-8.

[21] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1-24.

[22] MALHOTRA P, RAMAKRISHNAN A, ANAND G. LSTM-based encoder-decoder for multi-sensor anomaly detection[EB/OL]. (2016-07-11)[2020-10-06]. https://arxiv.org/abs/1607.00148.

[23] BOYD S, VANDENBERGHE L. Convex optimization[M].UK: Cambridge University Press, 2004.

［24］ BISHOP C M. Mixture density networks［M］. Birmingham, UK：Aston University, 1994.

Authors　Ms. **CHEN Lijing** received her B.S. degree in traffic and transportation from Jiangsu University, Zhenjiang, China, in 2018, and M.S. degree in applied traffic and transportation engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China in 2021. She is currently an engineer at Aviation Key Laboratory of Science and Technology on Fault Diagnosis and Health Management, Shanghai, China. Her research interest includes air traffic intelligence and aircraft abnormal behavior detection.

Dr. **ZENG Weili** received his B. S. degree in applied mathematics from Jishou University, Jishou, China, in 2007, M.S. degree in applied mathematics and Ph. D. degree in intelligent transportation systems from Southeast University, Nanjing, China, in 2009 and 2013, respectively. He is currently an associate professor and master supervisor with College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China. From 2013 to 2015, he was a postdoctoral researcher at School of Automation, Southeast University. His current research interest includes artificial intelligence, air traffic management, air intelligent transportation systems, and traffic data mining.

Author contributions　Ms. **CHEN Lijing** complied the models, conducted the analysis, interpreted the results and wrote the manuscript. Dr. **ZENG Weili** contributed to the model building and polished the manuscript. Dr. **YANG Zhao** contributed to the discussion and background of the study. All authors commented on the manuscript draft and approved the submission.

Competing interests　The authors declare no competing interests.

（Production Editor：ZHANG Bei）

一种基于深度混合密度网络的航空器轨迹异常检测方法

陈丽晶[1,2]，曾维理[1]，羊　钊[1]

（1.南京航空航天大学民航学院,南京 211106,中国；

2.故障诊断与健康管理航空科技重点实验室,上海 201601,中国）

摘要：及时、准确的异常航空器轨迹检测是提高飞行安全的关键。然而,目前用于异常检测的统计和机器学习方法不能很好地表征航空器轨迹的特征。飞行轨迹数据的高维性、异质性和时序性,导致异常检测精度较低。因此,本文提出了一种基于深度混合密度网络(Deep mixture density network,DMDN)的异常轨迹检测方法,可以对飞行轨迹安全性进行评估,并检测具有异常数据模式的飞行。该技术由两部分组成：利用深度长短时记忆(Long short-term memory,LSTM)网络对飞行轨迹特征进行有效编码;利用高斯混合模型(Gaussian mixture model,GMM)参数化飞行轨迹的统计特性。在广州白云国际机场终端区空域的实验结果表明,该方法能够有效地捕获航空器飞行轨迹的统计特性。该模型能够检测风险较高的异常航班,其性能优于两种主流方法,可作为空中交通管制的辅助决策工具。

关键词：航空器轨迹；异常检测；混合密度网络；长短期记忆；高斯混合模型