Article

Aircraft Track Anomaly Detection Based on MOD-Bi-LSTM

Yupeng Cao, Jiangwei Cao, Zhiguo Zhou * and Zhiwen Liu *

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China; yupengcao@bit.edu.cn (Y.C.); 3220190487@bit.edu.cn (J.C.)

* Correspondence: zhiguo Zhou (Z.Z.); zwliu@bit.edu.cn (Z.L.)

Abstract: In order to ensure flight safety and eliminate hidden dangers, it is very important to detect aircraft track anomalies, which include track deviations and track outliers. Many existing track anomaly detection methods cannot make full use of multidimensional information of the relevant track. Based on this problem, an aircraft track anomaly detection method based on the combination of the Multidimensional Outlier Descriptor (MOD) and the Bi-directional Long-Short Time Memory network (Bi-LSTM) is proposed in this paper. Firstly, track deviation detection is transformed into the track density classification problem, and then a multidimensional outlier descriptor is designed to detect track deviation. Secondly, track outliers detection is transformed into a prediction problem, and then a Bi-LSTM model is designed to detect track outliers. Experimental results based on real aircraft track data indicate that the accuracy of the proposed method is 96% and the recall rate is 97.36%. It can detect both track deviation and track outliers effectively.

Keywords: anomaly detection; Multidimensional Outlier Descriptor (MOD); Bi-directional Long-Short Time Memory (Bi-LSTM); track outliers; track deviation

Citation: Cao, Y.; Cao, J.; Zhou, Z.; Liu, Z. Aircraft Track Anomaly Detection Based on MOD-Bi-LSTM. *Electronics* **2021**, *10*, 1007. <https://doi.org/10.3390/electronics10091007>

Academic Editor: Tomasz Trzcinski

Received: 20 March 2021

Accepted: 19 April 2021

Published: 23 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid development of satellite communication technology, the ADS-B data reflecting aircraft track is becoming more abundant, and there is a lot of valuable information from the track data. In order to realize multi-source information fusion and the tracking of aircraft flights, Chinese and other scholars have carried out a lot of exploration and research work on track data mining methods, such as the ADS-B anomaly data detection model based on VAE-SVDD [1], track clustering [2], track correlation analysis [3,4], object motion pattern recognition [5], path planning [6], anomalous trajectories detection [7], etc. In the field of information fusion, the abnormal behavior of the target can be mined through the multi-dimensional track characteristics of the target, which is of great significance for situation assessment, threat assessment and command decision-making [8]. In the real environment, there are many unreasonable sampling points that have huge differences with their neighboring track points in motion features in the track data. These points are outliers in the track data [9]. All kinds of outliers affect the data quality of the track, and also have a certain effect on the information mining in subsequent tracks. Therefore, anomaly detection based on track data is an important basis of track data analysis and mining.

The track data of an aircraft flight is usually a sequence comprised of Multidimensional data points. The abnormal behavior of the target can be mined through the anomaly detection of Multidimensional tracks. Scholars at home and abroad have done a lot of research on anomaly detection and track anomaly detection. Chandola et al. summarized the anomaly detection technology, classification, application scope, advantages and disadvantages [10,11]. Existing anomaly detection methods can be divided into statistics-based anomaly detection [12], distance-based anomaly detection [13], density-based anomaly detection [14], depth-based anomaly detection [15] and deviation-based anomaly detection [16], etc. Existing track anomaly detection methods include methods based on the Multidimensional Local Outlier Factor [8], extracting global track features [17], classifiers [18,19], track segment similarity detection [20,21], etc. In recent years, with the

advancement of deep learning methods in the big data detection field, anomaly detection methods of deep learning have come to include an anomaly detection algorithm based on multi-layer convolution neural network interactive visualization [22], a Long-short Time Memory Network anomaly detection algorithm based on the encoder-decoder framework [23], a novel intrusion detector based on deep learning hybrid methods [24], forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques with the applications in supply chain management [25], a new method for anomaly detection of seismic preprecursor data based on LSTM-RNN [26], a deep learning-based hybrid intelligent intrusion detection system [27], deep learning for anomaly detection [28], a track anomaly detection algorithm based on the Bidirectional Long-short Time Memory Network [9], etc. In addition, Ruff et al. published a review on deep and shallow anomaly detection [29].

Some of the above methods can only detect the position anomaly of the target track, some do not make full use of the multidimensional features of the target, some rely too much on the parameter threshold, some are unable to automatically learn the difference between the abnormal points and normal points of the track and some have poor applicability to complex types of data. In a word, there are some limitations in mining the abnormal behavior of a target.

In view of the above problems, a track anomaly detection algorithm based on the Multidimensional Outlier Descriptor (MOD) combined with the Bi-directional Long-Short Time Memory Network (Bi-LSTM) is proposed in this paper. Firstly, the MOD algorithm is used to detect the track deviation. Then a 5-dimensional motion feature vector is built for each track point, and the feature vectors of track data within a time interval are selected as the input of the Bi-LSTM network to detect various abnormal behaviors of an aircraft.

2. Algorithm Description

2.1. Track Anomaly Detection Based on MOD-Bi-LSTM

The aircraft track anomaly includes the track deviation and the track outliers. A track deviation means the flight of the aircraft deviates from the fixed flight route. The track outliers refers to the abnormal height, speed and heading of aircraft at some track points. The flowchart of the algorithm for track anomaly detection based on MOD-Bi-LSTM is shown in Figure 1.

2.1.1. Data Preprocessing

Due to some reasons, the flight data of aircraft may be missing some of the time. In order to not affect the detection of track anomaly, the missing data need to be filled. According to the distribution of data, the mean and median can be used to fill the data. If the data is uniformly distributed, the mean is used to fill the data, and if the data is non-uniformly distributed, the median is used to fill the data.

2.1.2. Track Deviation Detection

The MOD algorithm is used to detect the track deviation. It is necessary to determine the threshold of the MOD algorithm. Since the MOD algorithm is to detect static data, the nearest neighbor number k and the anomaly threshold T can be determined through experimental verification. In the same scene, once the values of k and T are determined, these values can be adopted in the subsequent track deviation detection. The MOD value of the tracks is calculated, and tracks with a MOD value higher than the threshold are detected as a deviational track, then the deviational tracks will be removed.

Figure 1. Flowchart of MOD-Bi-LSTM anomaly detection algorithm.

2.1.3. Track Outliers Detection

- (1) Model training: The data is divided into training set and test set. The Bi-LSTM model uses the information of the current time, pre-t time and post-t time. The track features were extracted by forward and backward LSTM, respectively. In the feature extraction process, each track point is represented by a 5-dimensional vector: $x_t = \{\text{longitude, latitude, height, heading, speed}\}$. The x_0 represents the input information of the track point at the current time. The x_{-t} and x_t represent the track point information at the time t before and after x_0 , respectively. The input of the model is the sequences of track point vectors, and the features between the sequences are extracted automatically by Bi-LSTM. In the training process, the Bi-LSTM model can automatically learn the relationship between the track sequences.
- (2) Flight prediction: The test set is added into the trained Bi-LSTM neural network model to predict the flight time series. Various evaluation indexes are used to evaluate the algorithm performance, and the prediction results are compared with the LSTM and Bi-LSTM algorithm without MOD preprocessing.
- (3) Aircraft abnormal behavior detection: the predicted sequence value and the real sequence value are differentiated to obtain the difference sequence, and the difference sequence is solved by the SVDD algorithm to obtain the threshold of abnormal detection. In the test phase, whether the flight state of the aircraft is abnormal is judged by comparing the differentials with the threshold.

2.2. MOD for Track Deviation Detection

In anomaly detection, Local Outlier Factors (LOF) can not only solve the problem of anomaly detection with inconsistent local density [30] well, but also measure the position anomaly reflected by Euclidean distance well. For multidimensional target track data, LOF is not able to measure the position, speed and heading anomaly of the target for different requirements. Therefore, the MOD based on Dynamic Time Warping (DTW) algorithm can be the anomaly measurement of a multidimensional track. MOD can not only measure the position anomaly of the target, but also measure the height, speed and heading anomaly of the target, simultaneously.

Track data generally is multidimensional sequences comprised of multidimensional data points. In civic air traffic monitoring and processing systems, track data usually includes multidimensional characteristics such as the batch number, attribute, category, quantity, model, aircraft number, time, longitude, dimension, altitude, speed and heading of the target. Tracks can be expressed in following sets:

$$TD = \{TR_1, TR_2, \dots, TR_i, \dots, TR_n\} \quad (1)$$

where: TD is the track set, $i \in [1, n]$ is the track number and n is the total number of tracks; the track TR_i is a multidimensional track point sequence composed of several multidimensional track points in time sequence:

$$TR_i = \{P_{i1}, P_{i2}, P_{ij}, P_{im}\} \quad (2)$$

where: P_{ij} stands for the j th multidimensional track point in the i th track, $j \in [1, m]$ for the number of track points, and m for the total number of track points. For different tracks TR_i , m is not necessarily the same. P_{ij} for the track point is a vector with multidimensional characteristics, i.e., the target number, attribute, type, time, longitude, dimension, height, speed and heading of the j th multidimensional track point in the i th track. The DTW similarity function as the distance measurement replaces the traditional Euclidean distance to evaluate the distance between sequences to solve the problem that it is difficult to accurately calculate the distance between time series by using the Euclidean distance. The principle of DTW algorithm is to give the two sequences, $TR_A = \{TR_{A1}, TR_{A2}, \dots, TR_{Ai}, \dots, TR_{An}\}$ and $TR_B = \{TR_{B1}, TR_{B2}, \dots, TR_{Bj}, \dots, TR_{Bm}\}$, whose lengths are n and m . A matrix grid of $n \times m$ is constructed, with Euclidean distance D as the standard and matrix elements (i, j) as the distance $D(TR_{Ai}, TR_{Bj})$ between the point TR_{Ai} and the point TR_{Bj} . The matrix angle at the beginning of the sequence is the boundary condition, satisfying the constraints of continuity and monotony. The path with the smallest accumulative distance calculated by dynamic programming is the best path. At this point, the accumulative distance value of the best path is the DTW similarity of the two sequences. When calculating the DTW distance between the two tracks, only position characteristics are considered instead of the dynamic characteristics such as speed, heading and acceleration in terms of the Euclidean distance between the two points. The multi-factor directional DTW distance δ_M of the track TR_A and the track TR_B is defined by the multi-factor distance $mfdist(P_a, P_b)$ between two points. Considering the position, velocity, heading and acceleration features between the two vector points, the multi-factor distance between them is defined as follows:

$$mfdist(P_a, P_b) = w_d * dist(P_a, P_b) + w_v * dist(v_{P_a}, v_{P_b}) + w_\theta * dist(\theta_{P_a}, \theta_{P_b}) + w_\alpha * dist(\alpha_{P_a}, \alpha_{P_b}) \quad (3)$$

where: v_{P_a} and v_{P_b} is the speed of the point P_a and the point P_b ; θ_{P_a} and θ_{P_b} is respectively the heading of P_a and P_b ; α_{P_a} and α_{P_b} is respectively the heading of P_a and P_b . Euclidean Distance $dist(v_{P_a}, v_{P_b})$ is the Velocity Characteristic between P_a and P_b ; Euclidean distance $dist(\theta_{P_a}, \theta_{P_b})$ is the heading characteristic between P_a and P_b . w_d is the weight factor of the position feature; w_v is the weight factor of speed characteristics; w_θ is the weight factor of heading characteristics; w_α is the weight factor of acceleration character-

istics. The weight factor satisfies the formulation: $w_d \geq 0, w_v \geq 0, w_\theta \geq 0, w_\alpha \geq 0$, and $w_d + w_v + w_\theta + w_\alpha = 1$.

In the anomaly detection of multidimensional tracks, the values of four feature weights can be determined proportionally according to the needs of different detection tasks. When considering the position anomaly only, $w_d = 1, w_v = 0, w_\theta = 0, w_\alpha = 0$. Based on the defined multi-factor distance $mfdist(P_a, P_b)$, the dimension of directional DTW distance will be expanded, and the multi-factor directional DTW distance between multidimensional tracks TR_A and TR_B is defined as:

$$\delta_M(TR_A, TR_B) = mfdist(TR_{Ai}, TR_{Bj}) + \min \left\{ mfdist(TR_{A(i-1)}, TR_{B(j-1)}), mfdist(TR_{A(i-1)}, TR_{Bj}), mfdist(TR_{Ai}, TR_{B(j-1)}) \right\} \quad (4)$$

Among them: $i \in [1, n], j \in [1, m]$. $\delta_M(TR_A, TR_B)$ represents the similarity of multidimensional tracks TR_A and TR_B . Calculate MOD for multidimensional tracks $TR_i \in TD$:

- (1) Calculate the multi-factor neighbor boundary distance of the track:

$$\delta_{M \text{ border}}(TR_i, k) = \delta_M(TR_i, NN(TR_i, k)). \quad (5)$$

- (2) Calculate the neighboring track of the track TR_i :

$$\mathcal{N}(TR_i, k) = \{TR_j \in TD \setminus TR_i \mid \delta_M(TR_i, TR_j) \leq \delta_{M \text{ border}}(TR_i, k)\}. \quad (6)$$

- (3) Calculate the multi-factor reachable distance from the track TR_i to the track TR_j :

$$\delta_{M \text{ reach}}(TR_i, TR_j, k) = \max\{\delta_{M \text{ border}}(TR_i, k), \delta_M(TR_j, TR_i)\}. \quad (7)$$

- (4) Calculate the nearest neighbor density of the track TR_i :

$$\rho(TR_i, k) = \frac{|\mathcal{N}(TR_i, k)|}{\sum_{TR_j \in \mathcal{N}(TR_i, k)} \delta_{M \text{ reach}}(TR_i, TR_j, k)} \quad (8)$$

In the formula, replacing δ_M with $\delta_{M \text{ reach}}$ can smooth the neighbor density. The larger the value k , the smoother the effect becomes.

- (5) Calculate the Multidimensional Outlier Descriptor of the track TR_i :

$$MOD(TR_i, k) = \frac{\sum_{TR_j \in \mathcal{N}(TR_i, k)} \rho(TR_j, k)}{\rho(TR_i, k) |\mathcal{N}(TR_i, k)|} \quad (9)$$

where: the greater the ratio of the neighbor density of the tracks TR_i and the neighbor of the track TR_i , the greater the degree of anomaly of the track TR_i becomes.

2.3. Bi-LSTM for Track Outliers Detection

Recurrent neural network (RNN) is a recursive neural network whose input are sequences data. RNN has the characteristics of memory and parameter sharing, so it can learn the nonlinear features of sequences with high efficiency. Although RNN can deal with nonlinear time sequences effectively, it has the problem of gradient explosion and disappearance for time sequences with a too long delay. The Long Short Term Memory Network (LSTM) is an evolutionary network of traditional recurrent neural networks, which can improve the gradient explosion and disappearance problem. However, the predicted output of LSTM is determined by the input of previous multiple moments, and may lead to the loss of useful information when extracting data features. In many cases, the prediction is affected by the input of previous and subsequent multiple moments.

Therefore, this paper adopts the Bi-LSTM network with a forward and backward structure for track outliers detection [31].

2.3.1. Introduction of Bi-LSTM Neural Network Model

Bi-LSTM consists of a forward Long Short Term Memory Network (LSTM) and a backward LSTM. Because track outliers do not occur in isolation, the forward and backward structure of the Bi-LSTM model is more suitable for track outliers detection. The LSTM neural network unit is the basic unit that constitutes the Bi-LSTM neural network, and its network cell structure is shown in Figure 2.

Figure 2. LSTM structure.

The LSTM model can be regarded as an optimized recurrent neural network model, which is mainly aimed at the gradient problem in the process of long sequence training. The model is composed of input signal x_t at time t , cell state C_t , temporary cell state $\hat{C}_t$, hidden layer state h_t , forget gate f_t , memory gate i_t , and output gate o_t [32]. The working principle is as follows:

- (1) The forget gate screens weak correlation information and deletes it:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (10)$$

where h_{t-1} represents the hidden layer information at the previous moment; x_t represents the current input; W_f, b_f is the training parameter; σ is the sigmoid function (neural network activation function); f_t represents the weight of retained information.

- (2) The input gate screens the information of the strong correlation degree, and the sigmoid layer and the hidden layer jointly update the information in the cell state:

$$C_t = f_t \cdot C_{t-1} + \sigma(W_c[h_{t-1}, x_t] + b_c) \cdot \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (11)$$

where W_c and b_c represent training parameters; C_{t-1} represents the information of a previous cell state; C_t represents information about the current state of the cell.

- (3) The output gate determines the final output information:

$$h_t = \sigma(w_0[h_{t-1}, x_t] + b_0) \cdot \tanh(C_t) \quad (12)$$

where w_0 is the weight matrix, and the weight matrix used in this paper is a random matrix, that is, random initialization is carried out at the beginning of model training, and the weight matrix is iteratively updated through the back propagation of training loss. b_0 is the offset and σ is the activation function.

Combining forward LSTM and backward LSTM, a Bi-LSTM model can be obtained. The Bi-LSTM structure is shown in Figure 3.

Figure 3. Bi-LSTM structure.

Input signals $\{x_0, x_1, \dots, x_n\}$ are fed into the forward LSTM model and the backward LSTM model, respectively. Then output the forward implicit vectors $\{h_{F0}, h_{F1}, \dots, h_{Fn}\}$ and the backward implicit vectors $\{h_{B0}, h_{B1}, \dots, h_{Bn}\}$. The forward and backward implicit vectors are spliced to get: $\{[h_{F0}, h_{Bn}], [h_{F1}, h_{B(n-1)}], \dots, [h_{Fn}, h_{B0}]\}$, namely $\{h_0, h_1, \dots, h_n\}$.

2.3.2. Bi-LSTM Neural Network Model Construction

- (1) In flight data, each track point is represented by a 5-dimensional vector: $F_t = \{\text{longitude, latitude, height, heading, speed}\}$. Feature extraction is carried out on the data to form the training set of the model:

$$F_t = \{f_1, f_2, \dots, f_n\} \quad (13)$$

- (2) Data standardization processing; $X = \{X_1, X_2, \dots, X_n\}$ represents the processed training set, which is used as the model input, with $Y = \{Y_1, Y_2, \dots, Y_n\}$ representing model output.
- (3) The Bi-LSTM network is initialized; the initial weight matrix is assigned as a random number evenly distributed between $[0,1]$. The maximum number of iteration training is set as 256, the window size is 32, the learning rate is 0.001 and the minimum error value is 0.
- (4) Forward calculation: the hidden layer is input by the training set X and the Bi-LSTM model contains bidirectional LSTM hidden layers. Two groups of features are extracted during training. Each LSTM hidden layer contains n cell units arranged

in chronological order and the output through the hidden layer can be expressed as: $H_F = \{H_{F0}, H_{F1}, \dots, H_{Fn}\}$ and $H_B = \{H_{B0}, H_{B1}, \dots, H_{Bn}\}$.

- (5) Reverse error calculation: the characteristics are transferred from the LSTM in the forward and the backward motions and the results are calculated to obtain the difference loss between the real value and the predicted value. According to the loss, the whole network is backtracked and the parameters are modified. In order to improve the generalization ability of the model, a dropout mechanism is added between the Bi-LSTM layer and the full connection layer of the first layer to prevent overfitting of the model [33].
- (6) When the training times and error values meet the set requirements, the model training is stopped and the test set is predicted.

2.3.3. SVDD Evaluates Threshold

To solve the adaptive problem of detection threshold of aircraft abnormal behavior, the difference sequence could be obtained by taking the difference between the predicted sequence value in this paper and the real sequence value. Based on the difference sequence, a Support Vector Domain Description (SVDD) classifier [34,35] is achieved to judge flight anomaly data. The main idea of SVDD classifier design is divided into several steps. Firstly, the difference sequence is mapped to high-dimensional space by nonlinear mapping. Then, the smallest hypersphere containing all or most of the difference sequence samples is found in the high-dimensional space. Finally, the obtained hypersphere functions as the discrimination boundary for anomaly detection (the discrimination rule is: if the sample point of the difference sequence falls into the hypersphere in the high-dimensional space, the sample point is discriminated as a normal point; if the sample point of the difference sequence falls outside the hypersphere in high-dimensional space, judge the sample point as an abnormal point).

Based on the difference sequence between the predicted sequence and the real sequence, the radius r and the center a of the hypersphere can be obtained by SVDD solution, thus obtaining the classifier.

3. Experimental Evaluation

3.1. Experimental Setup

In this paper, the MOD-Bi-LSTM model is constructed using TensorFlow framework, and the model is trained on NVIDIA Titan XP graphics card. Because the test data in this study contains 5 features, the input size of the MOD-Bi-LSTM model is set to 5. Additionally, the number of hidden units is 128, the dropout rate is 0.5, the batch size is 1000 and the epoch is set to 256. The ReLU activation function and Adam optimizer are used in the model, and the initial learning rate is set to 0.001.

3.2. Data Set

In this paper, the real aircraft track data is used to verify the performance of the proposed method. The flight data of CCA1315, CCA1369, CCA1883 and CCA1803 within a month are extracted in this paper, and each datum includes the information of flight time, height, speed, heading, longitude and latitude. However, the distribution of positive and negative samples is unbalanced in the data set. In this paper, the data set is randomly divided into a training set and test set. The combination of an undersampling method and oversampling method is used to balance the category distribution of training data. The Synthetic Minority Oversampling Technique (SMOTE) is a kind of oversampling method. Its main idea is to form new minority samples by interpolating between several minority samples. However, SMOTE will produce noise samples during the interpolation process. This problem can be solved by using the Edited Nearest Neighbor (ENN) method to clean up the interpolation results. Any sample that is different from its k nearest neighbor categories will be removed, thus generating a class-balanced training set.

3.3. Loss Function

The loss function is used to measure the inconsistency between the predicted value $f(x)$ and the real value y of the model. It is usually expressed by $L(y, f(x))$ and can also be called the Cost Function. The smaller the value of loss function, the better the fitting of the model.

The loss function used in this experiment is Mean Absolute Error (MAE), also known as L1-loss. MAE is the sum of absolute values of the difference between the target value and the predicted value, which can be used to measure the difference between the predicted value and the real value. The calculation formula is as follows:

$$L(y_i - \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

where y_i is the target value and $\hat{y}_i$ is the predicted value.

3.4. Evaluation Metrics

A variety of commonly used machine learning evaluation metrics are adopted, including Accuracy, Precision, Recall, F1-score, Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

For binary classification problem, the sample can be divided into a true positive example (TP), false positive example (FP), true negative example (TN) and false negative example (FN) according to the combination of real situation and model prediction results. The confusion matrix is shown in Table 1.

Table 1. Confusion Matrix.

Actual Class	Predicted Class	
	Abnormal	Normal
Abnormal	TP	FN
Normal	FP	TN

Accuracy is defined as the ratio of correctly predicted sample number to total sample number; precision is defined as the ratio of correctly predicted positive samples to the total predicted positive samples. The recall rate is defined as the ratio of correctly predicted positive samples to the actual total number of positive samples; The F1-score is defined as the harmonic mean of precision and recall rate. The calculation formulas of these three indicators are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (15)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

The root mean square error (RMSE) is the arithmetic square root of the square of the difference between the predicted value and the real value. When the predicted value is completely consistent with the real value, RMSE is equal to 0. The greater the error, the greater the RMSE value. The calculation formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

where y_i is the real value and $\hat{y}_i$ is the predicted value, with n indicating the number of values.

The Mean Absolute Percentage Error (MAPE) is a percentage value. The smaller the MAPE value, the better the accuracy of the prediction model. The calculation formula is as follows:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (20)$$

where y_i is the real value and $\hat{y}_i$ is the predicted value, with n indicating the number of values.

3.5. Analysis of Experimental Results

3.5.1. Track Deviation Detection

The MOD algorithm can be used to preprocess the track data and detect the track deviation from the route in the data. Take longitude, latitude and height for example, Figure 4 shows the track deviation detection result based on MOD, in which the red dotted line indicates the track deviated from the route and the green solid line indicates the normal track. Figure 4a–d respectively show the flight track of flights CCA1315, CCA1369, CCA1883 and CCA1803 within a month. According to the detection results, the tracks that obviously deviate from the fixed flight route are removed from the data set.

Figure 4. Track Deviation Detection Result.

In order to verify the effectiveness of the MOD algorithm, the MOD algorithm was compared with the LOF algorithm, and the results are shown in Table 2.

Table 2. The Performance of MOD and LOF.

Method	Accuracy/%	Precision/%	Recall/%	F1-Score/%	Test Time/s
LOF	93.97	97.11	95.87	96.49	0.82
MOD	97.26	99.04	97.78	98.41	0.92

3.5.2. Track Outliers Detection

After the track data which deviates from the route in the data set is removed, the remaining data is divided into a training set and test set, which take up respectively 70% and 30% of the set, and the training set is added into the Bi-LSTM network for training. The trained model predicts the flight information of the aircraft in the test set. The predicted sequence values are compared with the real sequence values. The prediction results of height, speed and heading are shown in Figures 5–7.

Figure 5. The Prediction of Height.

Figure 6. The Prediction of Speed.

Figure 7. The Prediction of Heading.

In order to verify the effectiveness and innovation of the proposed method, we can compare the model in this paper with the LSTM model, Bi-LSTM model, BP model, LR model and CNN model. The results are shown in Table 3.

Table 3. Comparison of Prediction Indicators of Different Models.

Model	Height		Speed		Heading	
	RMSE /(m)	MAPE /%	RMSE /(Km/h)	MAPE /%	RMSE /(degrees)	MAPE /%
LSTM	1249.43	16.32	21.80	2.79	21.92	11.36
Bi-LSTM	810.62	10.33	19.36	2.58	19.51	10.07
BP	1032.59	13.49	21.87	2.84	20.66	10.71
LR	1520.18	19.86	26.41	3.44	24.73	12.81
CNN	402.16	5.25	16.20	2.11	13.49	6.99
ours	279.35	3.76	15.45	2.01	13.08	6.78

To solve the adaptive problem of the threshold of aircraft anomaly detection, the SVDD classifier is obtained based on the difference sequence between the predicted sequence and the real sequence, and the automatic detection of track outliers is realized. Next, 100 tracks' data is randomly extracted from the data set. The flight speed in a certain period of time is changed to 0.5 times of the original speed, the flight height is increased by 50 m, and the heading is reduced by 10 degrees. The modified track data and normal track data are added into the SVDD classifier for testing. Out of the 100 abnormal test samples, a total of 96 test samples showed that the distance between the predicted value and the true value to the center of the hypersphere is greater than the threshold value r , so the detection accuracy is 96%. Out of 265 normal test samples, a total of 258 test samples showed that the difference between the predicted value and the true value is less than the threshold r from the center of the hypersphere, thus the recall rate of detection is 97.36%. Table 4 presents a comparison of the detection effects of the models in this paper (the LSTM model, Bi-LSTM model, BP model, LR model and CNN model) on track outliers of aircraft.

Table 4. Comparison of Testing Effects of Different Models.

Model	Accuracy/%	Recall/%	Test Time/s
LSTM	91.62	92.61	0.94
Bi-LSTM	95.51	94.28	1.37
BP	93.49	94.10	2.23
LR	90.98	86.92	0.98
CNN	95.93	90.21	4.07
ours	96.00	97.36	2.06

As shown in Table 4, the accuracy and recall of the algorithm in this paper are both higher than those of other methods, in particular the recall rate is much higher than that of other methods. A high recall rate means that the model can detect more real outliers, which is very important in anomaly detection.

4. Conclusions

In this paper, an aircraft track anomaly detection method based on the combination of MOD and Bi-LSTM is proposed, and the effectiveness of the method is verified by using the real aircraft track data. By analyzing the experimental results, the proposed method can detect both track deviation and track outliers well. The main conclusions of this article are as follows:

- (1) The track deviation detection problem is transformed into the track density classification problem, and the MOD is designed to detect track deviation. The accuracy and recall of the MOD algorithm are improved compared to the LOF algorithm.

- (2) The track outliers detection problem is transformed into the prediction problem, and the Bi-LSTM model is used to detect the track outliers. Compared with the traditional methods, the accuracy is improved.
- (3) The anomaly detection algorithm considers not only the track density information, but also the features of the track point. More comprehensive factors are taken into account and the accuracy has improved.

In future studies, more attention should be paid to the application of the model, and the response time of the model should be reduced to ensure real time detection. Next, some optimization algorithms will be used to further optimize the parameters of the MOD-Bi-LSTM model to improve the accuracy of anomaly detection.

Author Contributions: Conceptualization, Y.C., Z.Z. and Z.L.; methodology, Z.Z. and Z.L.; software, J.C.; formal analysis, Z.Z.; investigation, Y.C., J.C., Z.Z. and Z.L.; writing—original draft preparation, Y.C.; writing—review and editing, J.C., Z.Z. and Z.L.; supervision, Z.Z. and Z.L.; project administration, Z.L.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Equipment Pre-research Field Fund Thirteen Five-year (No. 61403120109).

Data Availability Statement: Data was obtained from the third party and are available from the authors with the permission of the third party.

Acknowledgments: We would like to thank Zhenjun Zhang and Yalin Huang for the beneficial discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Luo, P.; Wang, B.; Li, T.; Tian, J. ADS-B anomaly data detection model based on VAE-SVDD. *Comput. Secur.* **2021**, *104*, 102213. [\[CrossRef\]](#)
2. Lee, J.; Han, J.; Whang, K. Trajectory clustering: A partition-and-group framework. In Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, Beijing, China, 12–14 June 2007; pp. 593–604.
3. Li, W.; Huang, C. Receiver optimal track Analysis based on Angle and Doppler frequency for external emitter location system. *J. Radar.* **2014**, *3*, 660–665.
4. Qi, L.; Wang, H.; Liu, Y. An interrupt track paired correlation algorithm based on statistical double threshold. *Acta Radarica Sin.* **2015**, *4*, 301–308.
5. Zheng, Y.; Liu, L.; Wang, L.; Xie, X. Learning transportation mode from raw gps data for geographic applications on the web. In Proceedings of the 17th International Conference on World Wide Web, Beijing, China, 21–25 April 2008; pp. 247–256.
6. Bao, J.; Zheng, Y.; Mokbel, M.F. Location-based and preference-aware recommendation using sparse geo-social networking data. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 6–9 November 2012; pp. 199–208.
7. Zhang, D.; Li, N.; Zhou, Z.; Chen, C. iBAT: Detecting anomalous taxi trajectories from GPS traces. In Proceedings of the 13th International Conference on Ubiquitous Computing, Beijing, China, 17–21 September 2011; pp. 99–108.
8. Pan, X.; Wang, H.; He, Y.; Xiong, W.; Zhou, W. Abnormal behavior detection method based on multi-dimensional track characteristics. *Acta Aerophenica Sin.* **2017**, *38*, 254–263.
9. Han, Z.-R.; Huang, T.; Ren, W.; Xu, G. Detection algorithm of Trajectory Outliers based on Bi-LSTM model. *J. Radar* **2019**, *8*, 36–43.
10. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv.* **2009**, *41*, 1–58. [\[CrossRef\]](#)
11. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection for Discrete Sequences: A Survey. *IEEE Trans. Knowl. Data Eng.* **2012**, *24*, 1–16. [\[CrossRef\]](#)
12. Beniger, J.R.; Barnett, V.; Lewis, T.; Bendre, S.M. *Outliers in Statistical Data*; Wiley: New York, NY, USA, 1994; pp. 24–26.
13. Knorr, E.M.; Ng, R.T.; Tucakov, V. Distance-based outliers: Algorithms and applications. *VLDB* **2000**, *8*, 237–253. [\[CrossRef\]](#)
14. Jin, W.; Tung, A.K.; Han, J.; Wang, W. Ranking outliers using symmetric neighborhood relationship. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Singapore, 9–12 April 2006; pp. 577–593.
15. Struyf, A.; Rousseeuw, P.J. High-dimensional computation of the deepest location. *Comput. Stat. Data Anal.* **2000**, *34*, 415–426. [\[CrossRef\]](#)
16. Sarawagi, S.; Agrawal, R.; Megiddo, N. *Discovery-Driven Exploration of OLAP Data Cubes*; Springer: Berlin, Germany, 1998; pp. 67–69.
17. Knorr, E.M.; Ng, R.T. Algorithms for Mining Distance-Based Outliers in Large Datasets. *VLDB* **1998**, *98*, 392–403.
18. Li, X.; Han, J.; Kim, S. Motion-alert: Automatic anomaly detection in massive moving objects. In Proceedings of the International Conference on Intelligence and Security Informatics, San Diego, CA, USA, 23–24 May 2006; pp. 166–177.

19. Li, X.; Han, J.; Kim, S.; Gonzalez, H. ROAM: Rule- and Motif-Based Anomaly Detection in Massive Moving Object Data Sets. In Proceedings of the 2007 SIAM International Conference on Data Mining, Minneapolis, MN, USA, 26–28 April 2007; pp. 273–284.
20. Lee, J.; Han, J.; Li, X. Trajectory Outlier Detection: A Partition-and-Detect Framework. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Cancun, Mexico, 7–12 April 2008; pp. 140–149.
21. Liu, L.; Qiao, S.; Liu, B.; Le, J. Efficient Abnormal trajectory Detection algorithm based on R-Tree. *J. Softw.* **2009**, *20*, 2426–2435.
22. Bo, L.; Hong, Z. Pavement anomaly detection algorithm based on multi-channel wavelet convolutional neural network. *J. Cent. China Norm. Univ.* **2019**, *53*, 200–206.
23. Batista, G.E.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor.* **2004**, *6*, 20–29. [[CrossRef](#)]
24. Wang, S.; Xia, C.; Wang, T. A Novel Intrusion Detector Based on Deep Learning Hybrid Methods. In Proceedings of the 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Washington, DC, USA, 27–29 May 2019.
25. Nguyen, H.D.; Tran, K.P.; Thomassey, S.; Hamad, M. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *Int. J. Inf. Manag.* **2020**, *57*, 102282. [[CrossRef](#)]
26. Cai, Y.; Yue, T.; Yun, T.; Xing, H. A New Method for Anomaly Detection of Seismic Precursor Data Based on LSTM-RNN. *Appl. Geophys.* **2019**, *16*, 257–266. [[CrossRef](#)]
27. Ashfaq Khan, M.; Kim, Y. Deep Learning-Based Hybrid Intelligent Intrusion Detection System. *Comput. Mater. Contin.* **2021**, *680*, 671–687. [[CrossRef](#)]
28. Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep Learning for Anomaly Detection. *ACM Comput. Surv.* **2021**, *54*, 1–38. [[CrossRef](#)]
29. Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Müller, C.R. A Unifying Review of Deep and Shallow Anomaly Detection. *Proc. IEEE* **2021**, 1–40. [[CrossRef](#)]
30. Janssens, J.H.M.; Postma, E.O. One-class classification with LOF and LOCI: An empirical comparison. In Proceedings of the 18th Annual Belgian-Dutch Conference on Machine Learning, Tilburg, The Netherlands, 18–19 September 2009; pp. 56–64.
31. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
32. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [[CrossRef](#)]
33. Srivastava, N.; Hinton, G.; Krizhevsky, A. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
34. Zhao, Y.-P.; Xie, Y.-L.; Ye, Z.-F. A new dynamic radius SVDD for fault detection of aircraft engine. *Eng. Appl. Artif. Intell.* **2021**, *100*, 104177. [[CrossRef](#)]
35. Tao, X.; Chen, W.; Li, X.; Zhang, X.; Li, Y.; Guo, J. The ensemble of density-sensitive SVDD classifier based on maximum soft margin for imbalanced datasets. *Knowl. Based Syst.* **2021**, *219*, 106897. [[CrossRef](#)]