Advances in Computer and Communications, 2023, 4(3), 191-195
https://www.hillpublisher.com/journals/acc/
ISSN Online: 2767-2875

Aviation Safety QAR Data Mining and Statistical Analysis

Yu Zhan, Junchao Zhang, Cailian Luo, Li Hu, Yulu Song*

Guilin University of Technology at Nanning, Guangxi, China.

How to cite this paper: Yu Zhan, Junchao Zhang, Cailian Luo, Li Hu, Yulu Song. (2023) Aviation Safety QAR Data Mining and Statistical Analysis. *Advances in Computer and Communication*, 4(3), 191-195.
DOI: 10.26855/acc.2023.06.017

Received: May 30, 2023
Accepted: June 28, 2023
Published: July 24, 2023

***Corresponding author:** Yulu Song, Guilin University of Technology at Nanning, Guangxi, China.

Abstract

Flight safety is the fundamental guarantee for the survival and development of civil aviation transport industry. Serious flight accidents will not only bring huge economic losses to airlines, but also pose a great threat to the life safety of passengers. Therefore, we need to pay close attention to flight safety. Firstly, the pre-analysis is carried out, and it is found that the quantity directly related to the aircraft manoeuvring is the disc quantity and the rod quantity, which correspond to roll manoeuvring and pitch manoeuvring respectively. After that, cluster analysis and graph neural network model introduced attention mechanism were used to train the data, and it was concluded that the most important operational factors affecting the flight of aircraft were rod size and attitude. According to the model, the reason for the heavy landing is traced back: "It is because of the incorrect loose rod that the rod size and attitude are abnormal, which makes the G-value curve of the landing significantly convex". Finally, the advantages and disadvantages of the model are evaluated and the sensitivity is analyzed. The model established in this paper can be extended to evaluate the stability of social and economic operation.

Keywords

Graph neural network, Aviation safety decision tree, Regression tree

1. Introduction

Flight safety is the basis of survival and development of civil aviation transportation industry. With the rapid development of civil aviation industry, it is more and more important to study the safety of flight. On March 21, 2022, the "March 21" crash ended China's civil aviation safety record of 10,590,000 flight hours. The occurrence of serious flight accidents will not only bring huge economic losses to the airlines, but also cause great threat to the lives of the passengers. Therefore, it is necessary to focus on flight safety issues, strengthen aviation safety research, comprehensively utilize existing data to strengthen scientific management, effectively improve the quality of employees through targeted and systematic control means, monitor and warn risks, and thus reduce the probability of flight accidents.

In terms of specific research and application of flight quality monitoring, the research of Chinese civil aviation industry can be divided into two aspects at present. One is to exceed the event research, analysis and application; The second is the statistical analysis and application of non-transfinite data. For the study of over-limit events, the over-limit threshold is generally set in the specified concentrated area of flight parameters, and the flight records exceeding the threshold are found out for key analysis, so as to prevent potential hazards from causing serious flight accidents. At present, this kind of analysis is the main body of flight quality monitoring work, which can better ensure the safety work at the present stage, but its deficiency lies in the lack of analysis of the reason of overlimit [1].

During the whole flight process from take-off to landing, the aircraft ensures flight safety through a series of flight controls, including roll control, pitch control and so on. At present, domestic airlines monitor flight controls through the over-limit, which can quickly identify aircraft status deviations, but can only tell safety managers what is happening, but not immediately find out the reason for such deviations. To this end, the reason for this deviation can be analyzed through the process change of the joystick.

2. Model establishment and solution

Inspired by the advantages of regression tree algorithm, the graph attention network model is considered. Graph attention network is a machine learning model for processing graph data. It can learn the feature vectors of nodes and edges to represent the structure and content of the graph, and use these feature vectors to perform tasks such as node classification, edge prediction and graph classification, which can improve the representation ability of the graph data, reduce the number of training parameters, so as to reduce the information redundancy between features and form the combination between features. Moreover, compared with convolutional neural network, it has the advantage of directly giving the comparison of convergent results, which is very suitable for rational quantitative description of aircraft maneuvering.

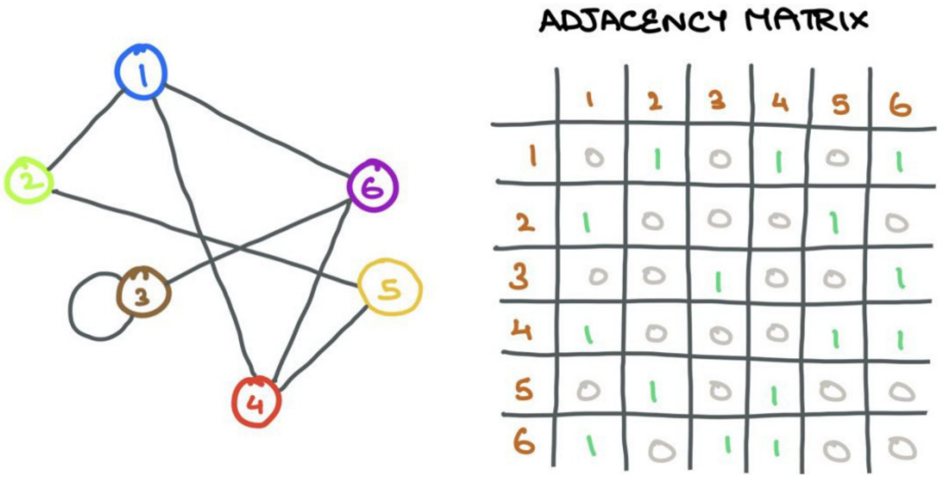

Figure 1. Structure and adjacency matrix.

Suppose that the graph convolution method in the single layer of graph convolution neural network is written as the following formula [2]:

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

The symbol description in the formula is as follows:

(1) $H^{(l)} \in \mathbb{R}^{N\times D}$ is the surface network input, $H^{(0)} = X$ is the initial input, the number of nodes in the graph is N, and each node is D represented by the dimensional feature vector [3];

(2) $\widetilde{A} = A + I_N$ indicates that a self-connecting adjacency matrix is added, and $\widetilde{D}$ is the degree matrix, satisfying the formula as follows:

$$\widetilde{D_{ii}} = \sum_j \widetilde{A_{ij}} \tag{2}$$

(3)The parameter matrix with training can be denoted as: $W^l \in \mathbb{R}^{D\times D}$

(4) The activation function is represented by σ.

Since the definition of convolutional layer is known, the whole neural network can be set up. In general, the subject graph convolutional neural network is composed of three convolutional layer [4], and the mathematical expression of the whole graph convolutional neural network is as follows:

$$Z = f(X, A) = soft\max\left(\hat{A}\mathrm{Re}LU\left(\hat{A}\mathrm{Re}LU\left(\hat{A}XW^{(0)}\right)W^{(1)}\right)W^{(2)}\right) \tag{3}$$

In Formula (3), there are the following stipulations: X represents the input characteristic matrix $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$, and the weight matrices of the three convolution layers are denoted as $W^{(0)}$, $W^{(1)}$ and $W^{(2)}$ respectively [5].

2.1 Calculation process of attention layer in Figure

In view of the fundamental meaning of GAT to guide the graph neural network into the attention mechanism, the following focuses on the calculation process of the graph attention layer:

Suppose N is the number of nodes, F is the number of node features, F' is the dimension of a new node feature vector, h is a matrix of $N \times F$ size [6], and R only represents the features of a node.

The input of the graph's attention layer can be written in the form of a set of node feature vectors, namely:

$$h = \left\{\vec{h}_1, \vec{h}_2, \cdots, \vec{h}_N\right\}, \vec{h}_i \in \mathbb{R}^F \tag{4}$$

Similarly, the output of the graph's attention layer is a new node feature vector set, namely:

$$h' = \left\{\vec{h'}_1, \vec{h'}_2, \cdots, \vec{h'}_N\right\}, \vec{h'}_i \in \mathbb{R}^{F'} \tag{5}$$

In the next step, the attention mechanism is implemented for each node. Suppose $a(.)$ is a function [7]. Table i, j below represents the mathematical expression of the attention coefficient of the *i* th and *j* th nodes:

$$e_{ij} = a\left(W\vec{h}_i, W\vec{h}_j\right) \tag{6}$$

Regularization of all adjacent nodes is introduced, and the calculation formula is:

$$a_{ij} = soft\max{}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \tag{7}$$

Combined with formula (6) and (7), the complete attention mechanism can be obtained after sorting and simplification. The mathematical expression is as follows:

$$a_{ij} = \frac{\exp\left(Leaky\,\mathrm{Re}LU\left(\vec{a}^T\left[W\vec{h}_i \| W\vec{h}_j\right]\right)\right)}{\sum_{k \in N_i} \exp\left(Leaky\,\mathrm{Re}LU\left(\vec{a}^T\left[W\vec{h}_i \| W\vec{h}_k\right]\right)\right)} \tag{8}$$

In Formula (8), $||$ is the connection operation, T is the transpose, and a_{ij} is the attention coefficient after normalization based on e_{ij} [8]. The attention coefficient between different nodes after regularization can be obtained through the above operations, which can be used to predict the output characteristics of each node, as shown in Equation (9):

$$\vec{h}'_j = \sigma\left(\sum_{j \in N_i} a_{ij} W \vec{h}_j\right) \tag{9}$$

In Formula (9), W is the weight matrix multiplied by features, *a* is the number of attention interrelations obtained by previous calculation, σ is the nonlinear activation function, and the traversal in $j \in N_i$ represents all nodes adjacent to i. If there are K independent attention mechanisms, the extended multi-head attention mechanism is adopted to achieve a stable self-attention learning process, k-average is adopted to replace the joining operation, and the application of the final nonlinear function is delayed. The final formula is shown in Equation (10):

$$\vec{h}'_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i} a_{ij}^k W^k \vec{h}_j\right) \tag{10}$$

In Equation (10), the first attention mechanism is, and the linear transformation weight matrix of input features under the first attention mechanism is expressed as.

2.2 Cluster analysis of data

Through the quantitative description of the above aspects, we can better understand the improper control of the pilot in the process of heavy landing, so as to provide a basis for improving the flight operation. Therefore, data of landing G value, attitude, lever change, wind direction, wind speed and throttle lever position with large weights were selected for analysis. Since the lever is recorded 5 times per second, this operation affects attitude (pitch Angle). Firstly, the relationship between the lever change and attitude change of the joystick is studied to quantify the pilot's operating intensity. If the operating intensity at a certain moment is significantly higher than that at other normal moments, it indicates that there is a problem in the flight process of the aircraft at that moment.

Firstly, cluster analysis is used to analyze some selected data:

1) Select the number of clustering k

k is the number of clusters, which usually need to be selected according to the characteristics of the actual problem. If the data distribution is not too obvious, you can use regular rules of thumb, such as the elbow rule or contour coefficient, to estimate the optimal k value.

2) Initialize the clustering center

k points are randomly selected from the data as clustering centers. Different initialization methods can be used, such as the k-Means ++ or Forgy methods.

3) Distribute data points to the nearest clustering center

For each data point, calculate its distance from each cluster center and assign it to the nearest cluster center. The Euclidean distance or Manhattan distance is usually used as a distance metric.

4) Update the clustering center

For each cluster, the average of all its member points is calculated and this average is used as the new cluster center.

5) Repeat steps 3 and 4 until convergence

Repeat steps 3 and 4 until the clustering center does not change or a preset number of iterations is reached.

6) Output clustering results

Output the cluster label to which each data point belongs.

The cluster analysis of the two features found that the rod weight and attitude data appeared abnormal values in similar time periods, indicating that there may be special circumstances in the flight process of the aircraft.

3. Trace the cause of aircraft deviation

Next, draw a line chart and make a comparative analysis on the data of rod size and attitude in the order of time, because both of them are recorded five times in one second. In order to draw a change image of rod size and attitude data, it is necessary to preprocess the data, that is, add the first data recorded in the last second to the fifth data in the last second. Simply, the matrix [M,N] is reshaped into a [1, MN] matrix, and the data are preprocessed with Matlab software.

From the above analysis of aircraft control, we can draw the conclusion that during the flight, due to the abnormal changes in wind speed and direction, the control range of the aircraft increases, which has a great impact on the rod weight. Especially in the time period indicated by the arrow, the rod weight, the tilt Angle and the landing G value all change synchronously. It is precisely because of the incorrect release of the rod. The landing G value is abnormal.

4. Conclusion

Decision tree algorithm can generate an easy to interpret tree structure, people can easily understand the meaning of each node in the tree, and thus have higher confidence in the reliability of the model. Decision tree algorithms can handle many types of data, including classification and regression problems. In addition, it can be used to process data with multiple categories or labels. Decision tree algorithm has good robustness and can deal with missing data and noisy data. Compared with the traditional graph convolutional neural network, the graph convolutional neural network with attention mechanism can greatly reduce the amount of computation and improve the efficiency of training and reasoning while maintaining high accuracy. Adding an attention mechanism can make the model pay more attention to important nodes and edges, thus making the model more robust to changes in input data. Since the attention mechanism can provide the weight of importance for each node and edge, the graph convolutional neural network with the addition of attention mechanism has strong interpretability.

5. Discussion

The Quick Access Recorder (QAR) data mining and statistical analysis of aviation safety is an important research area that can help airlines and relevant agencies better understand flight operations and flight safety. Future research directions may include the following:

Advanced data mining technology: With the development of technology, more and more data mining technology has been applied to the analysis of aviation safety QAR data. Future research directions could explore more advanced machine learning, artificial intelligence and data mining algorithms to improve the mining effect and accuracy of data.

Multi-source data integration analysis: In addition to QAR data, aviation safety research can also be combined with other data sources, such as airline operation data, meteorological data, air traffic control data, etc. Future studies could explore how to integrate multi-source data for comprehensive analysis to obtain a more comprehensive understanding of aviation safety.

Risk assessment and prediction: Using QAR data for risk assessment and prediction is one of the important applications of aviation safety. Future research could explore how to use QAR data for more accurate and reliable risk assessment and prediction to help airlines take corresponding measures to improve flight safety.

Air accident investigation support: QAR data plays an important role in air accident investigation. Future research could focus on how to better use QAR data to support air accident investigations by in-depth analysis of the information in QAR data to reveal the causes of accidents and potential improvement measures.

Visualization and user interface design: The results of analysis of QAR data should be presented visually to relevant decision makers and operators. Future research could focus on how to design visual interfaces that are intuitive and easy to understand, enabling airlines and relevant agencies to better understand and utilize the results of QAR data analysis.

In conclusion, the future research direction of aviation safety QAR data mining and statistical analysis will involve more advanced data mining technology, multi-source data integration analysis, risk assessment and prediction, aviation accident investigation support, visualization and user interface design and other aspects. These research directions will contribute to the improvement of aviation safety and airline flight operations.

References

[1] Li, B., Huang, J., Wang, X., Shi, Y., Lv, M., & Hu, M. (2022, August). Overview of Application in Data Mining Techniques to QAR Data Ansys. In 2022 2nd International Conference on Big Data Engineering and Education (BDEE) (pp. 13-21). IEEE.

[2] Wang, X., Zhao, X., & Yu, L. (2020, October). Data Mining on the Flight Quality of an Airline based on QAR Big Data. In 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT (pp. 955-958). IEEE.

[3] Zixuan, W., Weizhen, L., Cong, C., Qingchi, Y., & Hao, Z. H. A. N. G. (2020, October). Research on Influencing Factors of Fuel Flow Based on QAR Data. In 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT (pp. 800-804). IEEE.

[4] Jiao, Y., Han, J., Xu, B., Xiao, M., Shen, B., & Sun, H. (2021, October). Research on Domain Entity Extraction in Civil Aviation Safety. In 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT) (pp. 384-388). IEEE.

[5] Hong, W., Xiuxia, H., & Hongwei, W. (2008, December). Research and implementation of QAR data warehouse. In 2008 Second International Symposium on Intelligent Information Technology Application (Vol. 3, pp. 156-162). IEEE.

[6] Kang, Z., Shang, J., Feng, Y., Zheng, L., Liu, D., Qiang, B., & Wei, R. (2020). A deep sequence-to-sequence method for aircraft landing speed prediction based on QAR data. In Web Information Systems Engineering–WISE 2020: 21st International Conference, Amsterdam, The Netherlands, October 20-24, 2020, Proceedings, Part II 21 (pp. 516-530). Springer International Publishing.

[7] Han, J., & Jiao, Y. (2018, December). Conception of QAR Data Application Mode Based on Virtual Integration. In 2018 IEEE International Conference of Safety Produce Informatization (IICSPI) (pp. 392-395). IEEE.

[8] Li, L., Hansman, R. J., Palacios, R., & Welsch, R. (2016). Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring. Transportation Research Part C: Emerging Technologies, 64, 45-57.