

## 航空安全 QAR 数据挖掘与统计分析

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

桂林理工大学 广西南宁

**【摘要】**飞行安全是民航运输业生存发展的根本保障。严重的飞行事故不仅会给航空公司带来巨大的经济损失，也会对旅客的生命安全构成极大威胁。因此，飞行安全亟待引起重视。首先，进行预分析，发现与飞机机动直接相关的量是盘量和杆量，分别对应滚转机动和俯仰机动。之后，采用聚类分析和引入注意力机制的图神经网络模型对数据进行训练，得出影响飞机飞行最重要的运行因素是杆量和姿态。根据模型追溯重着陆的原因：“是由于不正确的松杆操作，导致杆尺寸和姿态异常，使着陆 G 值曲线明显上凸”。最后，评估了模型的优缺点并进行了灵敏度分析。本文建立的模型可以扩展到评估社会经济运行的稳定性。

**【关键词】**图神经网络；航空安全决策树；回归树

**【收稿日期】**2025 年 6 月 3 日

**【出刊日期】**2025 年 7 月 9 日

**【DOI】**10.12208/j.ae.20250004

### Aviation safety QAR data mining and statistical analysis

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

Guilin University of Technology, Nanning, Guangxi

**【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 简介

飞行安全是民航运输业生存发展的基础。随着民航业的快速发展，飞行安全研究愈发重要。2022 年 3 月 21 日，“3·21”空难事件终结了中国民航 1059 万飞行小时的安全纪录。重大飞行事故的发生不仅会给航空公司带来巨大的经济损失，也会对旅客的生命安全造成极大威胁。因此，需要关注飞行安全问题，加强航空安全研究，综合利用现有数据加强

科学管理，通过有针对性、系统性的管控手段有效提升从业人员素质，监测预警风险，从而降低飞行事故发生的概率。

在飞行品质监控的具体研究与应用方面，目前中国民航业的研究主要可以分为两个方面，一是超限事件的研究、分析与应用；二是非超限数据的统计分析与应用。对于超限事件的研究，一般在指定的飞行参数集中区域内设定超限阈值，找出超过阈

\*通讯作者：Yulu Song

注：本文于 2023 年发表在 Advances in Computer and Communications 期刊 4 卷 3 期，此文为其授权翻译版本。

值的飞行记录进行重点分析,防止潜在隐患引发严重飞行事故。目前,此类分析是飞行品质监控工作的主体,能够较好地保障现阶段的安全工作,但其不足之处在于缺乏对超限原因的分析<sup>[1]</sup>。

通过一系列飞行控制动作来保障飞行安全,包括滚转控制、俯仰控制等。目前,国内航空公司通过超限监控飞行控制,可以快速识别飞机状态偏差,但只能告诉安全管理人员正在发生什么,而无法立即找出偏差的原因。为此,可以通过操纵杆的过程变化来分析这种偏差的原因。

## 2 模型建立与求解

受回归树算法优点的启发,考虑图注意力网络模型。图注意力网络是一种处理图数据的机器学习模型,它可以学习节点和边的特征向量来表示图的结构和内容,并利用这些特征向量执行节点分类、边预测和图分类等任务,从而提高图数据的表示能力,减少训练参数的数量,从而减少特征间的信息冗余并形成特征之间的组合。而且,与卷积神经网络相比,它具有直接给出收敛结果对比的优势,非常适合对飞机机动进行合理的定量描述。

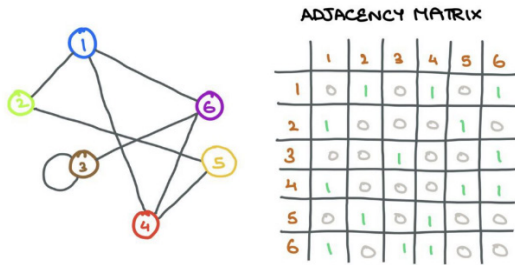


图1 结构和邻接矩阵

假设图卷积神经网络单层中的图卷积方法写为如下公式<sup>[2]</sup>:

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

公式中符号解释如下:

(1)  $H^{(l)} \in \mathbb{R}^{N \times D}$  是表面网络输入,  $H^{(0)} = X$  为初始输入, 图中节点数为  $N$ , 每个节点用  $D$  维特征向量表示<sup>[3]</sup>;

(2)  $\tilde{A} = A + I_N$  表示添加一个自连接邻接矩阵,  $\tilde{D}$  为度矩阵, 满足如下公式:

$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij} \quad (2)$$

(3) 训练后的参数矩阵可以表示为:  $W^l \in \mathbb{R}^{D \times D}$

(4) 激活函数表示为  $\sigma$ 。

知道了卷积层的定义,就可以建立整个神经网络。一般来说,本课题的图卷积神经网络由三个卷积层<sup>[4]</sup>组成,整个图卷积神经网络的数学表达式如下:

$$Z = f(X, A) = \text{soft max}\left(\hat{A} \text{ReLU}\left(\hat{A} \text{ReLU}\left(\hat{A} X W^{(0)}\right) W^{(1)}\right) W^{(2)}\right) \quad (3)$$

式(3)中,有如下规定:  $X$  表示输入特征矩阵  $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ , 3 个卷积层的权重矩阵分别记为  $W^{(0)}$ 、 $W^{(1)}$  和  $W^{(2)}$ <sup>[5]</sup>。

### 2.1 图中 attention 层的计算过程

鉴于 GAT 的根本意义在于引导图神经网络引入注意力机制,下面重点介绍图注意力层的计算过程:

设  $N$  为节点数,  $F$  为节点特征数,  $F'$  为新节点特征向量的维数,  $h$  为  $N \times F$  大小为<sup>[6]</sup>的矩阵,  $R$  仅表示一个节点的特征。

图的注意力层的输入可以写成一组节点特征向量的形式,即:

$$h = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}, \vec{h}_i \in \mathbb{R}^F \quad (4)$$

类似地,图的注意力层的输出是一个新的节点特征向量集,即:

$$h' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}, \vec{h}'_i \in \mathbb{R}^{F'} \quad (5)$$

接下来,为每个节点实现注意力机制。假设  $a(\cdot)$  是一个函数<sup>[7]</sup>。下表  $i$  表示  $j$  第  $i$  个节点的注意力系数的数学表达式  $th$  和  $j$  第  $j$  个节点:

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

引入所有相邻节点的正则化,计算公式为:

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

结合公式(6)和公式(7),经过排序和简化后,可以得到完整的注意力机制,其数学表达式如下:

$$a_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a^T \left[W \vec{h}_i \| W \vec{h}_j\right]\right)\right)}{\sum_{k \in N_i} \exp\left(\text{LeakyReLU}\left(a^T \left[W \vec{h}_i \| W \vec{h}_k\right]\right)\right)} \quad (8)$$

式(8)中,  $\|$  为连接操作,  $T$  为转置操作,  $a_{ij}$  为基于  $e_{ij}$ <sup>[8]</sup>归一化后的注意力系数。通过以上操作

可以得到经正则化后不同节点之间的注意力系数,进而可用于预测各节点的输出特性,如式(9)所示:

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

式(9)中,  $W$  为特征乘以权重矩阵,  $a$  为之前计算得到的注意力关联个数,  $\sigma$  是非线性激活函数, 中的遍历  $j \in N_i$  表示与  $i$  相邻的所有节点。如果有  $K$  个独立的注意力机制, 则扩展的多头采用注意力机制实现稳定的自注意力学习过程, 采用 **k-average** 代替 **joining** 操作, 并对最终非线性函数的应用进行延迟, 最终公式如公式(10)所示:

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

在式(10)中, 第一个注意力机制为, 第一个注意力机制下输入特征的线性变换权重矩阵表示为。

## 2.2 数据聚类分析

通过对以上几个方面的量化描述, 可以更好地了解飞行员在重着陆过程中操控不当的情况, 为改进飞行操作提供依据。因此, 选取着陆  $G$  值、姿态、杆位变化、风向、风速以及负重较大的油门杆位置等数据进行分析。由于每秒记录 5 次杆位操作, 因此该操作会影响姿态(俯仰角)。首先研究操纵杆的杆位变化与姿态变化之间的关系, 量化飞行员的操作强度, 如果某一时刻的操作强度明显高于其他正常时刻的操作强度, 则表明该时刻飞机的飞行过程存在问题。

首先, 利用聚类分析对一些选定的数据进行分析:

### (1) 选择聚类数 $k$

$k$  是聚类数, 通常需要根据实际问题的特点来选择。如果数据分布不是太明显, 可以采用常规的经验法则, 例如肘部法则或轮廓系数, 来估算最优的  $k$  值。

### (2) 初始化聚类中心

从数据中随机选择  $k$  个点作为聚类中心。可以使用不同的初始化方法, 例如 **k-Means++** 或 **Forgy** 方法。

### (3) 将数据点分布到最近的聚类中心

对于每个数据点, 计算其与每个聚类中心的距离, 并将其分配给最近的聚类中心。通常使用欧几

里得距离或曼哈顿距离作为距离度量。

### (4) 更新聚类中心

对于每个聚类, 计算其所有成员点的平均值, 并将该平均值作为新的聚类中心。

重复步骤 3 和 4, 直到收敛

重复步骤 3 和 4, 直到聚类中心不再改变或者达到预设的迭代次数。

### (5) 输出聚类结果

输出每个数据点所属的聚类标签。

对两项特征进行聚类分析发现, 杆重和姿态数据在相似的时间段内出现了异常值, 这表明飞机在飞行过程中可能存在特殊情况。

## 3 追踪飞机偏差的原因

接下来, 绘制折线图, 按时间顺序对杆径和姿态的数据进行比较分析, 因为两者都在一秒钟内记录了五次。为了绘制杆径和姿态数据的变化图像, 需要对数据进行预处理, 即将最后一秒记录的第一个数据与最后一秒记录的第五个数据相加。简单来说, 将矩阵  $[M, N]$  重塑为  $[1, MN]$  矩阵, 并使用 **Matlab** 软件对数据进行预处理。

从以上对飞机操控的分析可以得出: 在飞行过程中, 由于风速、风向的异常变化, 飞机的控制范围增大, 对拉杆重量影响较大。尤其在箭头所示时间段, 拉杆重量、倾斜角度、落地  $G$  值均同步变化, 正是由于拉杆动作不正确, 导致落地  $G$  值异常。

## 4 结论

决策树算法可以生成易于解释的树形结构, 人们可以轻松理解树中每个节点的含义, 从而对模型的可靠性有更高的信心。决策树算法可以处理多种类型的数据, 包括分类和回归问题。此外, 它还可以用来处理具有多个类别或标签的数据。决策树算法具有良好的鲁棒性, 可以处理缺失数据和噪声数据。与传统的图卷积神经网络相比, 加入注意力机制的图卷积神经网络在保持较高准确率的同时, 可以大幅减少计算量, 提高训练和推理的效率。加入注意力机制可以使模型更加关注重要的节点和边, 从而使模型对输入数据的变化具有更强的鲁棒性。由于注意力机制可以为每个节点和边提供重要性的权重, 因此加入注意力机制的图卷积神经网络具有较强的可解释性。

## 5 讨论

快速访问记录器 (QAR) 数据挖掘和航空安全统计分析是一个重要的研究领域,可以帮助航空公司和相关机构更好地了解飞行运行和飞行安全。未来的研究方向可能包括以下几点:

**先进的数据挖掘技术:**随着科技的发展,越来越多的数据挖掘技术被应用于航空安全 QAR 数据的分析。未来的研究方向可以探索更先进的机器学习、人工智能和数据挖掘算法,以提高数据的挖掘效果和准确性。

**多源数据融合分析:**除了 QAR 数据,航空安全研究还可以结合其他数据源,例如航空公司运营数据、气象数据、空中交通管制数据等。未来的研究可以探索如何整合多源数据进行综合分析,以更全面地了解航空安全。

**风险评估与预测:**利用 QAR 数据进行风险评估与预测是航空安全的重要应用之一。未来的研究可以探索如何利用 QAR 数据进行更准确、更可靠的风险评估与预测,以帮助航空公司采取相应措施,提高飞行安全。

**航空事故调查支持:**QAR 数据在航空事故调查中发挥着重要作用。未来的研究可以重点关注如何更好地利用 QAR 数据支持航空事故调查,通过深入分析 QAR 数据中的信息来揭示事故原因并提出改进措施。

**可视化与用户界面设计:**QAR 数据分析结果应以可视化的方式呈现给相关决策者和运营商。未来的研究可以重点关注如何设计直观易懂的可视化界面,使航空公司和相关机构能够更好地理解和利用 QAR 数据分析结果。

综上所述,未来航空安全 QAR 数据挖掘与统计分析的研究方向将涉及更先进的数据挖掘技术、多源数据融合分析、风险评估与预测、航空事故调查支持、可视化与用户界面设计等方面,为提升航空安全水平和航空公司航班运行水平做出贡献。

### 参考文献

- [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.

**版权声明:** ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



**OPEN ACCESS**