

DOI:10.16356/j.1005-2615.2021.06.005

## 基于密度聚类与匹配算法的异常飞行行为挖掘

吴欣蓬<sup>1</sup>, 汤新民<sup>1</sup>, 毛继志<sup>2</sup>, 郭鸿滨<sup>2</sup>, 管祥民<sup>3</sup>

(1. 南京航空航天大学民航学院, 南京 211106; 2. 中国航空无线电电子研究所航空电子系统综合技术国防科技重点实验室, 上海 200241; 3. 中国民航管理干部学院民航通用航空运行重点实验室, 北京 100102)

**摘要:** 在空中交通愈加拥挤的背景下, 航空器的异常飞行行为的有效挖掘可以辅助管制员进行调配决策。现有方法只能辨识飞机空间位置特征异常, 存在水平可扩展性的局限。本文考虑位置、速度、高度和航向 4 个异常特征, 采用高度层划分策略、局部异常因子和快速覆盖树对基于密度的有噪声应用中的空间聚类(Density-based spatial clustering of applications with noise, DBSCAN)方法进行改进, 提出局部异常因子改进的考虑速度、方向及高度的基于密度聚类方法(Density-based spatial clustering considering speed, direction and high level improved by local outlier factor, LOFDBSC-SDH)密度聚类算法对正常航迹模式进行快速准确提取。然后, 基于正常航迹模式设计考虑过点时间和上述异常特征的航迹匹配算法, 挖掘异常飞行行为。最后, 通过实验仿真验证了本文方法的有效性和应用价值。

**关键词:** 密度聚类; 高度层划分; 快速覆盖树; 过点时间; 异常飞行行为

中图分类号: V328

文献标志码: A

文章编号: 1005-2615(2021)06-0863-09

## Abnormal Flight Behavior Mining Based on Density Clustering and Matching Algorithm

WU Xinpeng<sup>1</sup>, TANG Xinmin<sup>1</sup>, MAO Jizhi<sup>2</sup>, GUO Hongbin<sup>2</sup>, GUAN Xiangmin<sup>3</sup>

(1. College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China; 2. Key Laboratory of National Defense Science and Technology of Avionics System Integrated Technology, China Institute of Aeronautical Radio and Electronics, Shanghai 200241, China; 3. Key Laboratory of Civil Aviation General Aviation Operation, Civil Aviation Management Institute of China, Beijing 100102, China)

**Abstract:** As air traffic has become increasingly congested, the effective mining of abnormal flight behaviors can assist controllers to make decisions. Current methods can only identify abnormal spatial positions with limited horizontal scalability. This paper incorporates four abnormal characteristics, position, speed, altitude and heading, and improves the density-based spatial clustering of applications with noise (DBSCAN) by the altitude level division strategy, the local outlier factor and the faster cover tree. Then, the density clustering algorithm named density-based spatial clustering considering speed, direction and high level improved by local outlier factor (LOFDBSC-SDH) is proposed to effectively extract the normal track patterns of airplanes. Next, based on the normal track patterns, a track matching algorithm is designed to mine abnormal flight behavior patterns, including flights' cross-point time and above abnormal features. Finally, the effectiveness of the proposed scheme is verified by the experimental simulation.

**Key words:** density clustering; flight level division; faster cover tree; cross-point time; abnormal flight behavior

**基金项目:** 国家自然科学基金(61773202)资助项目; 中国航空无线电电子研究所航空电子系统综合技术国防科技重点实验室基金(6142505180407)资助项目; 中国民航管理干部学院民航通用航空运行重点实验室开放基金(CAMICKFJJ-2019-04)资助项目。

**收稿日期:** 2020-06-30; **修订日期:** 2021-01-21

**通信作者:** 汤新民, 男, 教授, 博士生导师, E-mail: tangxinmin@nuaa.edu.cn。

**引用格式:** 吴欣蓬, 汤新民, 毛继志, 等. 基于密度聚类与匹配算法的异常飞行行为挖掘[J]. 南京航空航天大学学报, 2021, 53(6): 863-871. WU Xinpeng, TANG Xinmin, MAO Jizhi, et al. Abnormal flight behavior mining based on density clustering and matching algorithm[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(6): 863-871.

伴随着航空运输全球化趋势,中国民航运输快速发展。中国民航2019年完成旅客运输6.6亿人次,较上年增长7.9%,其中国际航线完成的旅客运输量则高达7 425.1万人次,增长16.6%<sup>[1]</sup>。客运量的大幅增长给原本紧张的空域资源分配和调度带来了巨大压力,使得管制员和飞行员的工作负荷倍增。历年空管不安全事件报告显示,在这样的背景下,飞机的异常飞行行为日益频发,例如:因恶劣天气、管制员指挥不当、军民航冲突和无线电干扰等原因导致飞机在飞行过程中产生偏离既定航路,不满足最小间隔标准等异常飞行行为。

空中交通管制的主要任务是避免异常飞行行为的发生,保障飞机的安全运行。但是,目前在空管实践环节,自动化监视设备只能够在战术阶段对小于间隔的异常飞行行为进行预警,不能够发现上述多样化异常飞行行为,难以分析其机理并制定相关针对性措施。而在理论研究环节,随着新航行系统的发展,广播式自动相关监视(Automatic dependent surveillance-broadcast, ADS-B)技术在民航领域广泛推广,每天都会产生大量的航迹运行数据。许多学者采用航迹大数据对飞机异常飞行行为进行了探索性研究<sup>[2-7]</sup>。其中比较有代表性的是Ho等<sup>[8]</sup>基于ADS-B数据综合回归预测和鞅方法来辨识航迹空间位置不匹配数据的异常飞行行为,其异常偏航行为检出率达81%,但是无法辨识其他类型的异常飞行行为。Gariel等<sup>[9]</sup>以ADS-B数据中的飞行计划航路点和航迹转弯点来表征原始航迹,通过重采样方法和主成分分析法获取降维后的等长航迹序列,利用基于密度的有噪声应用中的空间聚类(Density-based spatial clustering of applications with noise, DBSCAN)算法的噪声辨识功能识别异常航迹,完成空域异常飞行行为的监视。王超等<sup>[10]</sup>对航迹空间特征利用层次聚类算法来辨识航迹样本类别,并进一步通过偏差建模来量化异常飞行的严重程度。潘新龙等<sup>[11]</sup>基于定义的多因素

定向Hausdorff距离,构造航迹多维度局部异常因子来辨识异常飞行行为,实现航迹的异常行为检测与挖掘,但是对多维特征采取欧氏距离度量存在误差,同时在邻居搜索时时间成本较高。Liu等<sup>[12]</sup>利用滤波算法剔除航迹数据的噪声和离群点,然后采用自然语言处理中的可变长度N-Grams算法来提取正常飞行行为特征向量,通过训练One-class SVM分类器来辨识异常飞行行为,但不适用于海量航迹的分析。

与上述研究方法不同的是,本文认为航迹是航空器在飞行员、管制员和客观环境因素影响下所表现出来的具有时空特性的行为轨迹。而航迹变化的形成机理主要是飞机速度、航向和高度调整的结果。同时,实际飞行航迹与航路存在偏差。因此,本文首先提出带速度、航向和高度层约束的局部异常因子改进的考虑速度、方向及高度的基于密度聚类方法(Density-based spatial clustering considering speed, direction and high level improved by local outlier factor, LOFDBSC-SDH),提取飞机正常航迹模式,并附加上相对时间特征;通过构建海量ADS-B航迹数据的快速覆盖树<sup>[13]</sup>来提高算法的处理速度;然后本文引入4D航迹中的过点时间<sup>[14]</sup>概念和过点时间的偏移约束开展实际航迹与正常航迹模式地相似性匹配;最终实现对航空器异常飞行行为地辨识挖掘。

## 1 基本概念

### 1.1 ADS-B航迹数据

ADS-B接收机异步解析航迹数据存在数据缺失问题,采用中值滤波算法进行填补;而对于航迹点极少的航班,即ADS-B有效接受范围边界处的航班数据,将视为无效数据进行剔除。最终获得包括飞机ICAO呼号-地址、时间、经度、纬度、高度、速度和航向等航迹信息,部分示意如表1所示。由此可见,飞机的历史航迹是一种时空大数据。

表1 ADS-B航迹数据

Table 1 ADS-B track data

ICAO呼号-地址	时间	对地速度/(m·s <sup>-1</sup> )	经度/(°)	纬度/(°)	真航向/(°)	高度/m
QTR871-06a1c0	2019/12/1 00:00:06:296	213.719 8	119.827 8	32.270 7	278.914 1	9 174.48
QTR871-06a1c0	2019/12/1 00:00:08:347	213.719 8	119.823 0	32.271 4	279.768 0	9 174.48

### 1.2 正常航迹模式

本文定义飞机正常航迹模式是指在历史无冲突表现的条件下,飞机在规定过点时间上位于既定的空间位置,并且速度、航向和高度特征与历史上的正常航迹模式相一致的类簇。因此,对飞机的航迹进行定义

$$\begin{cases} S = \{T_1, T_2, \dots, T_j, \dots, T_n\} \\ T_j = \{p_1^j, p_2^j, \dots, p_i^j, \dots, p_m^j\} \\ p_i^j = \{t_i^j, \lambda_i^j, \varphi_i^j, h_i^j, v_i^j, \theta_i^j\} \end{cases} \quad (1)$$

式中:S为飞机航迹数据集,其中第j个飞机的航迹数据集为 $T_j$ ,对于 $T_j$ 中第i个航迹点 $p_i^j$ 有经度 $\lambda_i^j$ 、纬度 $\varphi_i^j$ 、高度 $h_i^j$ 、对地速度 $v_i^j$ 和真航向 $\theta_i^j$ 对应过点时

间  $t_i^j$ 。过点时间  $t_i^j$  定义为相对于当天凌晨 0 点的毫秒计数,取值范围为 0~86 400 000 ms,该数据需要对表 1 时间数据按协调世界时(Universal time coordinated,UTC)进行转换。故同一正常航迹模式类簇可由一个 S 来表示。

### 1.3 异常飞行行为

所谓异常飞行行为,定义为飞机在运行过程中,表现出在过点时间、空间位置、速度、航向或高度层特征与历史正常航迹模式相异的飞行行为,即飞行表现违反 1.2 节正常航迹模式定义。因此,异常飞行行为与正常航迹模式是相反关系。

鉴于航迹需分段处理,进一步给出异常飞行行为在航迹分段条件下的定义:若一飞机航迹的任意子轨迹符合上述异常飞行行为定义,则称该飞机存在异常飞行行为。

## 2 基于 ADS-B 数据的正常航迹模式提取

### 2.2 正常航迹模式的提取原理

#### 2.2.1 约束的引入

异常飞行行为的形成过程主要表现为 2 种形式:(1)飞机之间在原本既定的高度层、航路上产生异常飞行行为,表现为同航路前后追赶、交叉航路相互接近。这主要是速度异常改变导致的。(2)飞机未在原本既定的高度层或航路上而产生异常飞行行为,表现为飞机偏航、穿越高度层而发生飞行冲突,这主要是航向和高度异常改变导致的。因此,正常航迹模式的提取应当将高度层、速度和航向作为必要特征。

传统 DBSCAN 轨迹挖掘算法<sup>[15]</sup>通过计算航迹点与其他航迹点之间的距离,将位于邻居半径  $\epsilon$  范围内的航迹点进行归并。如果归并结束后类簇的元素个数大于密度  $\mu$  则认为聚类成功;反之,将该类簇视为噪声剔除。该算法只考虑了航迹的空间位置关系。进一步作图 1 分析可知,虚线圆圈为  $\epsilon$  所形成的邻居半径范围,即水平间隔安全区域,而高度层下界高  $h_k$  和  $h_{k+1}$  刻画了垂直间隔标准。当航

迹  $T_i$  穿越高度层  $h_{k+1}$  后,其子轨迹  $T_{\text{sub}2}$  将不再与  $T_j$  发生聚类关系。因此,通过高度层标准来划分航迹可以规避不必要的聚类操作。目前,国内仅有交叉航路,不存在平行航路,因此交叉航路汇聚点处会出现不同类航迹点  $\epsilon$  邻域的交叠,即  $T_i$  和  $T_j$ ,这样将会引起错误的聚类。但是通过引入速度和航向特征可以很好地区分这 2 种航迹,从而将其标记为不同的类型。反之,对于同航路同航向且速度近似的航迹将被聚类为同一类簇,即  $T_m$  和  $T_n$ 。

此外,DBSCAN 基于欧氏距离判定航迹点之间的邻居关系。如果将这 3 个特征纳入欧氏距离计算会增加算法的运算时间。因此,为了兼顾速度、航向和高度层特征而又不增加欧氏距离计算复杂度,本文将 3 个特征以约束形式(式(2~4))引入 DBSCAN 算法进行改进,以使其适用于正常航迹模式的提取

$$h_k \leq h_{p_i}, h_{p_a} \leq h_{k+1} \quad (2)$$

式中  $\{h_k, h_{k+1}\} \subset H$ 。

$$|v_p - v_q| < \delta_v \quad (3)$$

$$|\theta_p - \theta_q| < \delta_\theta \quad (4)$$

式中: $H$  为高度层约束集合; $h_k$  为第  $k$  个飞行高度层的下边界高度值,因此对于满足高度层约束(2)的两航空器航迹点  $p_i$  和  $p_a$  将位于同一高度层中,可近似投影为  $h_k$  所在平面进行 DBSCAN 聚类,而无须在三维空间进行空间距离计算,节省计算时间。这即是“高度层划分策略”。速度和航向约束为式(3,4),其中  $\delta_v$  和  $\delta_\theta$  分别为速度阈值和航向阈值。可以依据飞机性能参数设定,从而辨识出相同飞行行为的航迹模式。此外,算法并未让式(1)中的时间特征  $t_i^j$  参与计算,而是将其作为附加属性,置于提取的正常航迹模式中。这主要是考虑到时间和空间度量尺不具有相同物理量纲,以及 ADS-B 数据的时间点由于异步解析可能无法对齐。因此,同一正常航迹模式中将存在时间不同而在空间上相似的轨迹。

#### 2.2.2 局部异常因子

虽然,在大部分时间下,飞机的航迹数据均属于正常航迹模式。但是过滤掉一些隐藏的异常模式,对本文借助模式匹配来辨识异常飞行行为是有必要的。原始的 DBSCAN 的异常点剔除能力有限,依赖于参数  $\epsilon$  和  $\mu$ 。为了获得较为干净的正常航迹模式,继续在 DBSCAN 的邻居判别过程中引入局部异常因子(Local outlier factor, LOF)<sup>[11]</sup>

$$\text{lrd}_\mu(p) = \frac{|\text{neighborPts}(p)|}{\sum_{q \in \text{neighborPts}(p)} \text{reachDist}_\mu(p, q)} \quad (5)$$

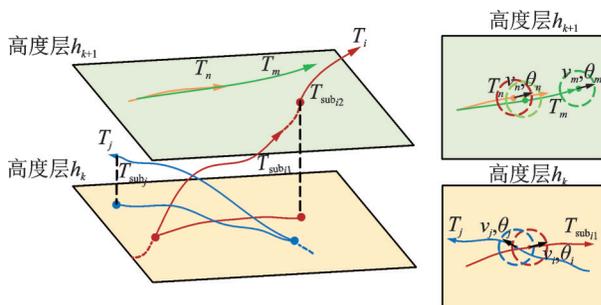


图 1 轨迹特征与正常航迹模式的联系  
Fig.1 Relationship between track characteristics and normal track patterns

$$\text{LOF}_\mu(p) = \left( \sum_{q \in \text{neighborPts}(p)} \text{lrd}_\mu(q) / \text{lrd}_\mu(p) \right) / |\text{neighborPts}(p)| \quad (6)$$

式中： $|\text{neighborPts}(p)|$ 为 $p$ 的邻居个数； $\text{reachDist}_\mu(p, q)$ 表示 $p$ 和 $p$ 的邻居 $q$ 之间在密度为 $\mu$ 条件下的可达距离； $\text{lrd}_\mu(p)$ 为航迹点 $p$ 的局部可达密度； $\text{LOF}_\mu(p)$ 为 $p$ 的局部异常因子。当 $\text{LOF}_\mu(p) \leq 1$ 时，则 $p$ 的局部可达密度与其邻居航迹点相接近，可视为属于同一类簇；否则， $p$ 为离群的异常点，进行剔除。

### 2.2.3 快速覆盖树

除此之外，DBSCAN和LOF的计算均以蛮力算法来寻找邻居对象，时间成本较高，不能满足航迹大数据的快速分析要求。本文提出构建ADS-B航迹数据集的快速覆盖树数据结构<sup>[13]</sup>来降低邻居查找过程的时间消耗。对一棵快速覆盖树而言，任意结点均包含一个单一的航迹点数据，并且满足3个不变量约束：

(1) 层次不变量即结点 $a$ 所对应的一个关联整数 $\text{level}(a)$ 。并且对结点 $a$ 的子结点 $b$ 存在如下等式约束关系

$$\text{level}(a) = \text{level}(b) - 1 \quad (7)$$

(2) 覆盖不变量定义为 $\text{covdist}(a) = 2^{\text{level}(a)}$ 。结点 $a$ 和其子结点 $b$ 的距离满足如下约束关系

$$d(a, b) \leq \text{covdist}(a) \quad (8)$$

(3) 分离不变量定义为 $\text{sepdist}(a) = 2^{\text{level}(a)-1}$ 。结点 $a$ 的任意两个子结点 $b_1, b_2$ 的距离满足如下约束关系

$$d(b_1, b_2) > \text{sepdist}(a) \quad (9)$$

依据上述结点与子结点约束关系可知，从根结点出发，每个父结点均向下覆盖其子结点，即子结点对应航迹点为其父结点对应航迹点的邻居。基于上述约束规则可以构建ADS-B航迹数据集对应的一棵快速覆盖树。

为了在快速覆盖树中查找航迹点 $p$ 的邻居，设结点 $a$ 的子结点集合为 $\text{children}(a)$ ，后代结点集合为 $\text{descendant}(a)$ ，定义结点 $a$ 到其后代结点之间的最大距离为

$$\text{maxdist}(a) = \underset{b \in \text{descendant}(a)}{\text{argmax}} d(a, b) \quad (10)$$

从根结点出发查找 $p$ 的邻居。分别向下按层次寻找 $p$ 所属的覆盖结点。对某层 $p$ 所属结点 $a$ ，计算该航迹点与结点 $a$ 的各子结点 $b$ 之间的距离，并按距离从小到大排序考察结点 $a$ 及其子结点 $b$ 和航迹点 $p$ 是否满足如下约束

$$d(a, p) > d(a, b) - \text{maxdist}(b) \quad (11)$$

若满足，则该子结点是 $p$ 在快速覆盖树下一层

次所属的覆盖结点。重复该步骤，直到所有子结点均不满足式(11)，则该父结点为 $p$ 的邻居。依据文献[13]可知， $n$ 结点快速覆盖树的查找时间复杂度为 $O(c^6 \log n)$ ，其中 $c$ 为常量，一般取2。故相对于使用蛮力算法查找邻居航迹点的幂级时间复杂度来说，使用快速覆盖树数据结构降低了算法查找邻居的时间成本。

### 2.2.4 LOFDBSC-SDH算法

将基于上述方案改进DBSCAN提取正常航迹模式的算法称为LOFDBSC-SDH，相关符号定义如表2所示，其伪代码如算法1所示。算法第2行首先基于 $\epsilon$ 邻居半径标准，按2.2.3节所述方法将输入的航迹点逐步插入到快速覆盖树相应结点，形成ADS-B航迹数据集对应的完整快速覆盖树，即建立算法的快速邻居搜索空间。算法第7行和18~28行在邻居搜索环节引入高度层约束 $[h_k, h_{k+1}]$ 、速度约束 $\delta_v$ 和航向约束 $\delta_\theta$ ，结合密度阈值 $\mu$ 完成同类航迹地辨识。算法8~11行基于上述邻居结果计算该航迹点的LOF值，辨识异常点。将异常点及其邻居一并剔除，节省算法聚类时间。待航迹数据中异常点剔除后，从第12~17行开始利用mergeClusters方法(34~41行)进行正常航迹类簇地合并，最终提取得到正常航迹模式。

表2 符号定义表

Table 2 Symbol definition

符号	定义
$S$	ADS-B航迹数据集
$H$	高度层约束集
$\delta_v$	速度阈值
$\delta_\theta$	航向阈值
$\mu$	密度阈值
$\epsilon$	邻居半径
$\text{neighborPts}(p)$	航迹点 $p$ 的邻居集合
$C$	正常航迹模式集合
FCT	ADS-B航迹数据的快速覆盖树
$\text{LOF}_\mu(p)$	$p$ 的局部异常因子
corePoint	类簇中心集合
$N_{\text{candidate}}$	候选邻居集合

### 算法1 LOFDBSC-SDH算法

Input:  $S, H, \epsilon, \mu, \delta_v, \delta_\theta$

Output:  $C$

(1) Procedure LOFDBSC-SDH( $S, \epsilon, \mu, \delta_v, \delta_\theta$ )

(2) Construct the Faster Cover Tree data structure of  $S$  based  $\epsilon$ : FCT;

```

(3) Initialize  $C = \{c_1, c_2, \dots, c_i, \dots, c_n\}$  and corePoint;
(4) For each flight point  $p$  in  $S$  do
(5)  $p = \{t, \lambda, \varphi, h, v, \theta\}$ ;
(6) Mark  $p$  as a classified flight trajectory point;
(7)  $\text{neighborPts}(p) = \text{acquireNeighbors}(p, \text{FCT}, \mu, H, \delta_v, \delta_\theta, \text{corePoint})$ ;
(8) Calculate  $\text{LOF}_\mu(p)$ ;
(9) If  $\text{LOF}_\mu(p) > 1$  then
(10) Delete  $\text{neighborPts}(p)$  from  $S$ ;
(11) Break;
(12) If  $|\text{neighborPts}(p)| \neq 0$  then
(13)  $C \leftarrow \text{neighborPts}(p)$ ;
(14) For  $c_i, c_j$  and  $c_i \neq c_j$  in  $C$  do
(15) If  $\text{mergeClusters}(c_i, c_j) == \text{TRUE}$  then
(16) Remove  $c_j$  from  $C$ ;
(17) Return  $C$ ;
(18) Procedure  $\text{acquireNeighbors}(p, \text{FCT}, \mu, H, \delta_v, \delta_\theta, \text{corePoint})$ :
(19) Define  $H = \{h_1, h_2, \dots, h_m\}$ ;
(20) Initialize  $\text{neighborPts}(p)$ ;
(21) Search FCT to acquire the candidate neighbors of  $p$ :  $N_{\text{candidate}}$ ;
(22) For  $q$  in  $N_{\text{candidate}}$  do
(23) If  $h_q, h_p \in [h_k, h_{k+1}]$  and  $\{h_k, h_{k+1}\} \subset H$  then
(24) If  $|v_p - v_q| < \delta_v$  and  $|\theta_p - \theta_q| < \delta_\theta$  then
(25)  $\text{neighborPts}(p) \leftarrow q$ ;
(26) If  $|\text{neighborPts}(p)| > \mu$  then
(27)  $\text{corePoint} \leftarrow p$ ;
(28) Return  $\text{neighborPts}(p)$ ;
(29) Procedure  $\text{mergeClusters}(c_i, c_j)$ :
(30) Initialize  $\text{flag} = \text{FALSE}$ ;
(31) For each  $q$  in  $c_j$  do
(32) If  $q \in \text{corePoint}$  then
(33)  $\text{flag} = \text{TRUE}$ ;
(34)  $c_i = c_i \cup c_j$ ;
(35) Break;
(36) Return  $\text{flag}$ .

```

结合 LOFDBSC-SDH 算法的设计思路和航空器实际运行场景,参数  $\epsilon$  可取最小水平间隔,而参数密度阈值  $\mu$  则一般依据经验设定<sup>[16]</sup>。

### 3 基于匹配算法的异常飞行行为辨识

异常飞行行为虽然有偏航和间隔缩小等具体形式,但是从宏观角度来看就是不同于正常航迹的异常模式。因此,基于匹配算法辨识实际航迹是否归属于正常航迹模式,将未匹配上的航迹认为存在异常飞行行为。

#### 3.1 航迹的分割

航迹中包含航路点、转弯起始点、转弯结束点等有效特征,常规对全航迹实施匹配的方法可能忽略掉这些局部特征。因此,基于航迹划分对其子航迹实施匹配以克服该问题,同时也可降低单次匹配的时间消耗。最小描述长度(Minimum description length, MDL)<sup>[17]</sup>能够为某种模式类中的所有成员寻找不可约、最小的特征表达方式。由于文献[17]提出的 MDL 是用于二维航迹划分,因此本文首先采用 2.2.1 节提出的高度层划分策略提取同高度层子航迹,然后将子航迹投影至高度层平面按 MDL 算法进行划分,获取该子航迹的有效特征表达。

#### 3.2 异常飞行行为辨识

基于航迹匹配辨识异常飞行行为需要采用合适的相似性度量。由于欧式距离 EucDistance 对航迹点的时间偏移敏感,动态时间规整距离时间复杂度高,最大公共子航迹距离对稀疏航迹度量效果较差,本文从航迹差异度的角度进行分析,采用 Hausdorff 距离<sup>[17]</sup>来实现航迹距离度量,并引入时间、速度和航向因素对其进行改进,以更好地判断异常飞行行为。

具体来讲,经过高度层划分和 MDL 算法分割提取到的子航迹  $T_i = \{p_1^{i,t_{i1}}, p_2^{i,t_{i2}}, \dots, p_m^{i,t_{im}}\}$  只需与同高度层的正常航迹模式  $T_j = \{p_1^{j,t_{j1}}, p_2^{j,t_{j2}}, \dots, p_n^{j,t_{jn}}\}$  进行匹配辨识,节省了计算时间。假设匹配比较的起始时刻  $t_{i1} = t_{j1}$ ,但由于 ADS-B 接收机异步解析或外界干扰因素,  $t_{sk}, t_{qk}$  不一定对应同一时刻。故算法 2 设计时间偏移阈值  $\delta_{TD}$  对时间差异进行容忍,以寻求近似同一时刻的数据。设航迹  $T_i$  和  $T_j$  分别对应时间范围  $[t_{i1}, t_{ik}]$  和  $[t_{j1}, t_{jk}]$ ,算法第 2 行和第 7~10 行首先按照时间先后次序遍历 2 段航迹的航迹点,当  $|t_{s,a} - t_{q,b}| \leq \delta_{TD}$  时,表示 2 个航迹点为同一时刻下的航迹点,可以进一步进行同一时空下的相似性度量,反之,则将取后一时间点做比较。

此外,在相似性度量上引入速度和航向约束,提出速度航向影响因子模型,如式(12)所示。通过  $f_{v\theta}$  来伸缩欧氏距离,如式(13)所示。当速度和航向偏差超过规定阈值时,即  $f_{v\theta} > 1$ ,  $f_{v\theta}$  使欧氏距离值成倍增长,达到相异区分的效果。反之,  $f_{v\theta} \leq 1$ ,即速度和航向均相似,此时航迹的相似性完全取决于各航迹点的空间位置关系 EucDistance。

$$f_{v\theta} = \begin{cases} \max\left(\frac{|v_a^{i,t_{sa}} - v_b^{j,t_{qb}}|}{\delta_v}, \frac{|\theta_a^{i,t_{sa}} - \theta_b^{j,t_{qb}}|}{\delta_\theta}\right) & \text{EucDistance} \leq \epsilon \\ 1 & \text{EucDistance} > \epsilon \end{cases} \quad (12)$$

进而得到式(13~14)构造两航迹 Hausdorff 距离计算式(15)

$$\|p_a^{i,t_{sa}} - p_b^{j,t_{sb}}\| = f_{\text{v0}} \cdot \text{EucDistance}(p_a^{i,t_{sa}}, p_b^{j,t_{sb}}) \quad (13)$$

$$\text{Hausdorff}(T_i, T_j) = \max_{p_a^i \in T_i} \min_{p_b^j \in T_j} \|p_a^i - p_b^j\| \quad (14)$$

$$\text{Hausdorff}(T_i, T_j) = \max(h(T_i, T_j), h(T_j, T_i)) \quad (15)$$

依据相似性匹配算法,如算法2所示,最终可分别在各高度层内完成子轨迹的相似性匹配,辨识异常飞行行为。

#### 算法2 航迹相似性匹配算法

Input: sub trajectories  $T_i = \{p_1^{i,t_{s1}}, p_2^{i,t_{s2}}, \dots, p_m^{i,t_{sm}}\}$  divided by height levels and MDL, normal trace pattern  $T_j = \{p_1^{j,t_{q1}}, p_2^{j,t_{q2}}, \dots, p_n^{j,t_{qn}}\}$

Output: Hausdorff distance similarity measure  $\text{Hausdorff}(T_i, T_j)$

- (1) For  $t_{s,a}, t_{q,b}$  in  $[t_{s1}, t_{sk}], [t_{q1}, t_{qk}]$  do:
- (2) If  $|t_{s,a} - t_{q,b}| \leq \delta_{\text{TD}}$  then
- (3) Return  $\text{Hausdorff}(T_i, T_j) = \max(\text{Hausdorff}(T_i, T_j), \text{Hausdorff}(T_j, T_i))$ ;
- (4) Else
- (5) If  $t_{s,a} \leq t_{q,b}$  then
- (6)  $t_{s,a} \rightarrow t_{s,a+1}$ ;
- (7) Else
- (8)  $t_{q,b} \rightarrow t_{q,b+1}$ ;
- (9) Return infinite.

结合航空器运行场景,  $\text{EucDistance}(p_a^{i,t_{sa}}, p_b^{j,t_{sb}})$  表征航空的间隔距离,因此不能超过  $\epsilon$ , 而  $f_{\text{v0}}$  的极限值为 1, 故匹配算法的后续步骤为: 当相似度量值  $\text{Hausdorff}(T_i, T_j) > \epsilon$  时, 实际航迹应视为存在异常飞行行为。反之, 视为正常。

## 4 实验分析

### 4.1 实验设计

本文选取华东地区某空域 ADS-B 接收机所获取的 2019-12-1 至 2019-12-30 清洗后的 1 776 207 条有效航迹数据记录作为训练数据集, 2019-12-31 清洗后 56 572 条有效航迹数据作为测试数据集。首先基于 LOFDBSC-SDH 算法利用训练数据提取正常航迹模式, 其中  $\epsilon$  取最小水平间隔 10 km<sup>[18]</sup>,  $\mu$  取 50, 高度层约束  $H$  设置参考文献[18], 依据速度航向变化值的统计分析,  $\delta_v$  取 100 m/s,  $\delta_\theta$  取 10°。然后, 基于  $\delta_{\text{TD}} = 2\ 000$  ms 运用匹配算法辨识测试数据集中的异常飞行行为。

最后, 为了验证本文提出异常飞行行为挖掘方案的准确性, 取清洗后的 2019-11-31 的 63 129 条有

效正常航迹(共 2 107 个航班)数据, 对其中 100 个航班的高度数据进行分段切分、混合后再拼接操作, 构造异常飞行行为航迹样本数据, 并给定标签 1; 然后继续各选取 100 个航班做速度、航向数据的相同处理, 其中航向异常数据, 即空间位置的偏差异常, 故其构造需要将航向数据连同经纬度数据一并更换, 分别给定标签 2 和 3, 最终生成 300 个异常飞行行为的航班航迹数据。其余未处理数据给定标签 0, 视为正常飞行航迹样。最终形成验证数据集。按照上述相同的异常飞行行为挖掘步骤处理验证数据集, 分析本文方法的准确率。

### 4.2 实验结果

#### 4.2.1 真实场景异常飞行行为挖掘

对 4.1 节所述训练数据集, 使用本文提出的 LOFDBSC-SDH 算法提取飞机正常航迹模式, 共获得 124 种正常航迹模式, 其可视化结果如图 2(a) 所示。从结果可知, LOFDBSC-SDH 算法相较于基线算法 DBSCAN, 如图 2(b), 所提取得到的正常航迹模式其轮廓更为清晰, 离群点较少, 方向性更强; 图 2(c) 左下角的许多航迹模式在 LOFDBSC-SDH 算法作用下得到强化(红色虚线区域), 同时图 2(b) 中心部分的橙色航迹表征出了该空域内的交叉航路事实, 具有典型的代表性。综上所述, 本文提出的 LOFDBSC-SDH 算法通过引入局部异常因子提高了算法 DBSCAN 的离群点剔除能力, 可以获取较为干净的正常航迹模式, 降低离群点对航迹匹配的干扰。

进一步采用 3.2 节异常飞行行为辨识方法对华东地区某研究空域某日全天航迹进行异常飞行行为挖掘, 结果如图 2(c) 所示。该航班直接从常熟市和张家港市上空飞跃, 而查阅航图数据发现并无此航路, 可以肯定存在异常飞行行为。

#### 4.2.2 实验结果分析

在正常航迹模式数据库中, 按时间节点和高度层约束进行检索, 分析数据, 发现图 2(b) 中的异常飞行行为主要与正常航迹模式 66 存在一定联系, 其所在时间、高度层、经度与模式 66 基本一致, 但是在纬度、航向和速度分布上存在较大差异, 如图 3 所示。该异常飞行行为的航班的飞行速度呈现加速状态, 与正常模式 66 所显示出的匀速变化状态相异。此外, 图 2(b) 显示器航迹不同于图 2(a) 中的各典型正常航迹模式, 即从常熟市上空直接汇入正常航迹所在航路, 该事实与图 3 所示与正常航迹模式存在显著纬度差异、23 200~23 400 ms 时间区间内存在显著航向差异的实验结果相符。

因此, 本文所提出的方案挖掘出了实际 ADS-B 数据中存在的违反速度和航向约束的异常飞行行为。

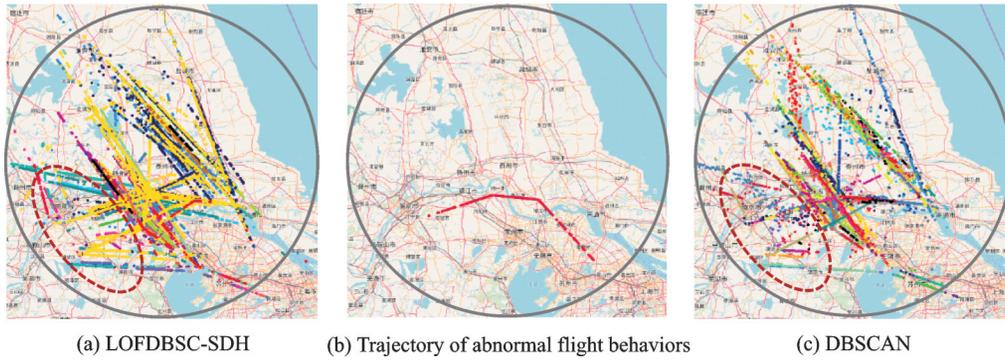


图 2 华东地区某研究空域

Fig.2 Research airspace in East China

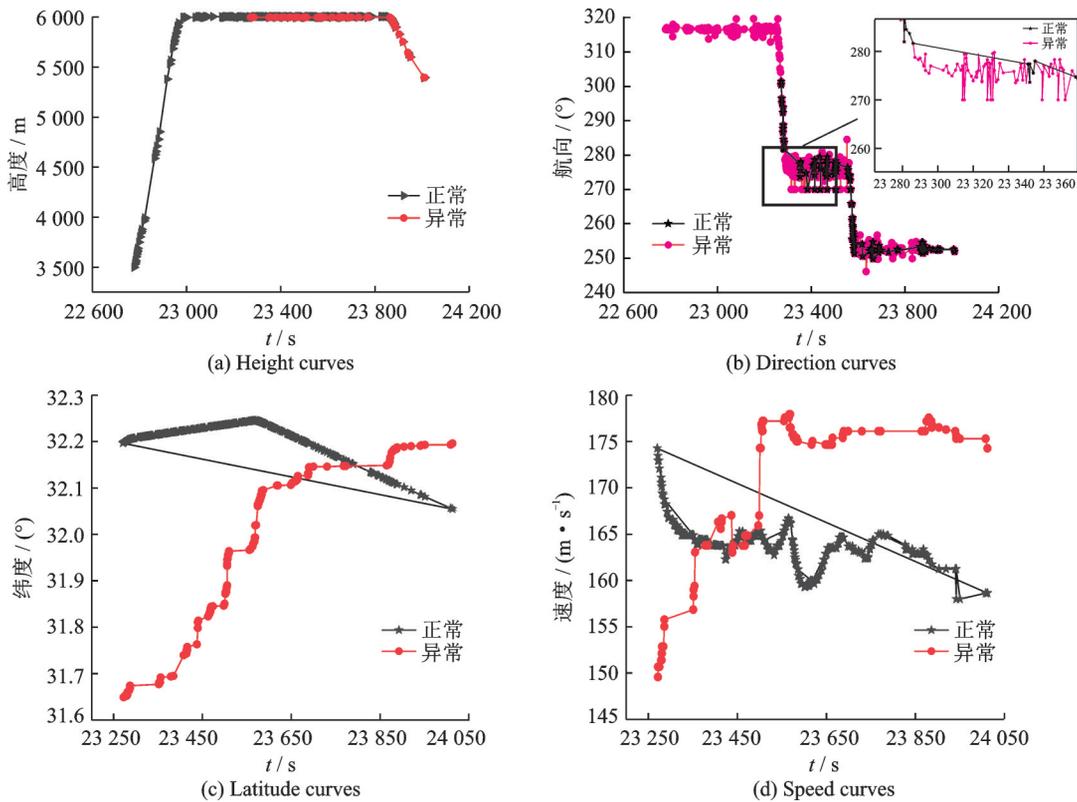


图 3 异常飞行行为特征曲线

Fig.3 Characteristic curves of abnormal flight behaviors

### 4.2.3 准确率分析

在验证数据集上,按 4.2.1 节相同实验步骤进行异常飞行行为地挖掘辨识。对有效结果进行统计分析,如表 3 所示。本文所提出的方案在验证数据集上能够辨识出其中 93.3% 的异常飞行行为,其中高度异常辨识准确度为 91%,速度异常辨识准确度为 96%,而航向异常辨识准确度为 93%。需要注意的是,位置异常是这 3 种异常的最终表

表 3 异常飞行行为挖掘准确性

Table 3 Mining accuracy of abnormal flight behaviors

指标	总体(包括位置异常)	高度异常	速度异常	航向异常
准确性	93.3	91.0	96.0	93.0

现,包含于总体准确性表述中。

综上所述,实验结果表明本文所提出的异常飞行行为挖掘方法对于实际飞机的异常飞行行为挖掘是有效的;并且,从异常飞行行为的产生机理出发,挖掘辨识飞机的过点时间、高度、速度和航向特征来判定飞机的异常状态,相较于引言中相关文献算法只通过空间位置特征偏差判定飞机异常飞行的做法,本文的方案更为合理。

### 4.3 算法分析

#### 4.3.1 聚类质量分析

对于没有基准的数据集,一般采用内在方法来评估聚类质量,DAVIES-BOULDIN 指标<sup>[16]</sup>,即 DBI,就是一种有效的方法,其定义如下

$$DBI = \frac{1}{N} \sum_{i=1}^N \max_{j \neq i} \left( \frac{\overline{\text{inDist}}_i + \overline{\text{inDist}}_j}{\|c_i - c_j\|_2} \right) \quad (16)$$

式中:对于 $N$ 个聚类结果, $\overline{\text{inDist}}_i$ 和 $\overline{\text{inDist}}_j$ 分别表示第 $i$ 和第 $j$ 个类别的内部平均距离, $c_i$ 和 $c_j$ 分别表示第 $i$ 和第 $j$ 个类别的聚类中心。因此,当类内部紧凑,类之间的间距较大时,聚类质量较好,所得DBI值较小。计算DBSCAN和LOFDBSC-SDH的DBI结果如表4所示。

表4 算法DBI比较

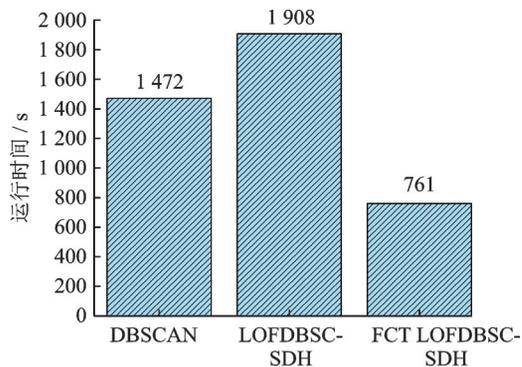
Table 4 Comparison of DBIs

算法	DBI指标
DBSCAN	0.775
LOFDBSC-SDH	0.683

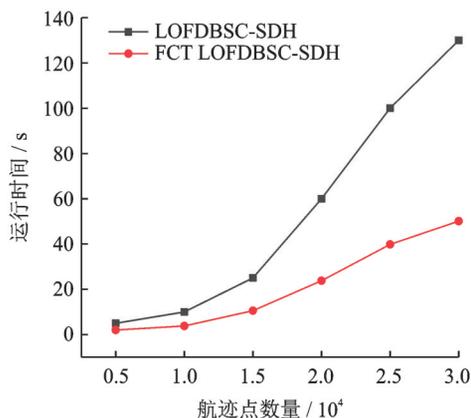
从表4结果可知,LOFDBSC-SDH算法相较于传统的DBSCAN算法具有更好的聚类效果,该评估结果与可视化结果图2(a,b)一致。

#### 4.3.2 算法运行效率分析

本文对引言中多数文献采用的基线算法DBSCAN、所提出的LOFDBSC-SDH算法以及引入快速覆盖树的FCT LOFDBSC-SDH进行了运行时间分析,其结果如图4所示。可见,LOFDBSC-SDH算



(a) Operation time



(b) Horizontal scalability

图4 算法性能比较

Fig.4 Comparison of algorithm performance

法引入LOF后,虽然在正常航迹模式的提取上取得了较好的效果,但是增大了算法的时间复杂度,运行时间有所增加。之后引入快速覆盖树数据结构,预先构建ADS-B航迹数据的邻居搜索空间,加快了算法邻居计算速度,降低了算法的时间复杂度,运行所需时间减少。因此,基于快速覆盖树的LOFDBSC-SDH算法在实验结果的优良性和算法的时间复杂度上做出一种良好地平衡,可以应用于海量ADS-B航迹数据的正常航迹模式提取,辅助异常飞行行为的辨识,规避引言相关方案<sup>[8,11-12]</sup>的不足。

## 5 结 论

本文提出了一套LOFDBSC-SDH密度聚类算法和匹配算法相结合的异常飞行行为挖掘方案。首先,为了克服传统方法只以飞机空间位置偏差作为异常飞行行为判定的不足,考虑异常飞行行为的产生原因,提出在位置异常基础上进一步考虑速度、高度和航向异常特征来设计挖掘算法。其次,为了弥补传统算法水平扩展局限性,提出高度层划分策略规避算法不必要聚类过程,并结合局部异常因子和快速覆盖树提出LOFDBSC-SDH算法,对海量ADS-B航迹数据进行正常航迹模式的快速、有效提取。然后,考虑过点时间和高度、速度、航向3个异常特征,设计相似度匹配算法来挖掘辨识异常飞行模式。最后,算法的DBI指标和运行时间实验表明,本文提出的LOFDBSC-SDH算法克服了传统DBSCAN的水平可扩展局限性,并提高了聚类的精度;而仿真实验结果表明所述方案能够有效辨识存在位置、速度、高度和航向异常的飞行行为,弥补了传统方法只能辨识空间位置偏差异常的不足;采用实际ADS-B数据的实验表明本文的方案能够挖掘真实运行场景中的异常飞行行为,具有良好地应用价值。

不过,本文方案还无法有效解决恶劣天气等环境因素所带来的影响等问题,需要在接下来的研究中引入多源数据进行进一步改进。

### 参考文献:

- [1] 中华人民共和国交通运输部. 2019年交通运输行业发展统计公报[EB/OL]. (2020-05-12) [2020-06-20]. [http://xxgk.mot.gov.cn/jigou/zhghs/202005/t20200512\\_3374322.html](http://xxgk.mot.gov.cn/jigou/zhghs/202005/t20200512_3374322.html).
- [2] 王超,韩邦村,王飞. 基于轨迹谱聚类的终端区盛行交通流识别方法[J]. 西南交通大学学报, 2014, 49(3): 546-552.  
WANG Chao, HAN Bangcun, WANG Fei. Identification of prevalent air traffic flow in terminal airspace based on trajectory spectral clustering[J]. Journal of

- Southwest Jiaotong University, 2014, 49 (3): 546-552.
- [3] 王莉莉,彭勃.基于LOFC时间窗分割算法的航迹聚类研究[J].南京航空航天大学学报,2018,50(5): 661-665.  
WANG Lili, PENG Bo. Track clustering based on lofc time window segmentation algorithm[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2018, 50(5): 661-665.
- [4] 徐涛,李永祥,吕宗平.基于航迹点法向距离的航迹聚类研究[J].系统工程与电子技术,2015,37(9): 2198-2204.  
XU Tao, LI Yongxiang, LYU Zongping. Research on track clustering based on normal distance of track point[J]. Journal of Systems Engineering and Electronics, 2015, 37(9): 2198-2204.
- [5] LI Lishuai, GABRIEL M, HANSMAN J R, et al. Anomaly detection in onboard-recorded flight data using cluster analysis[C]//Proceedings of 2011 IEEE/AIAA 30th Digital Avionics Systems Conference. Seattle, USA: IEEE, 2011: 4A4-1-4A4-11.
- [6] SABARISH B A, KARTHI R, KUMAR T G. Graph similarity-based hierarchical clustering of trajectory data[J]. Procedia Computer Science, 2020, 171: 32-41.
- [7] ZHONG Z, LEE E E, NEJAD M, et al. Influence of CAV clustering strategies on mixed traffic flow characteristics: An analysis of vehicle trajectory data[J]. Transportation Research Part C: Emerging Technologies, 2020, 115: 102611.
- [8] HO S, SCHOFIELD M, SUN B, et al. A Martingale-based approach for flight behavior anomaly detection[C]//Proceedings of the 20th IEEE International Conference on Mobile Data Management (MDM). Hong Kong, China: IEEE, 2019: 43-52.
- [9] GABRIEL M, SRIVASTAVA A N, FERON E. Trajectory clustering and an application to airspace monitoring[C]//Proceedings of IEEE Transactions on Intelligent Transportation Systems. [S. l.]: IEEE, 2011: 1511-1524.
- [10] 王超,徐肖豪,王飞.基于航迹聚类的终端区进场程序管制适用性分析[J].南京航空航天大学学报,2013, 45(1): 130-139.  
WANG Chao, XU Xiaohao, WANG Fei. ATC serviceability analysis of terminal arrival procedures using trajectory clustering[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2013, 45(1): 130-139.
- [11] 潘新龙,王海鹏,何友,等.基于多维航迹特征的异常行为检测方法[J].航空学报,2017,38(4): 254-263.  
PAN Xinlong, WANG Haipeng, HE You, et al. Anomalous behavior detection method based on multi-dimensional trajectory characteristics[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(4): 254-263.
- [12] LIU Yanfang, LYU Jianghua, MA Shilong. A real time anomaly detection method based on variable N-gram for flight data[C]//Proceedings of IEEE 20th International Conference on High Performance Computing and Communications (HPCC/SmartCity/DSS). Exeter, UK: IEEE, 2018: 00080.
- [13] IZBICKI M, SHELTON C. Faster cover trees[J]. European Journal of Operational Research, 2015, 184 (3): 1179-1181.
- [14] TANG Xinmin, CHEN Ping, ZHANG Yu. 4D trajectory estimation based on nominal flight profile extraction and airway meteorological forecast revision[J]. Aerospace Science and Technology, 2015, 45(1): 387-397.
- [15] 李宗林,罗可. DBSCAN算法中参数的自适应确定[J].计算机工程与应用,2016,52(3): 70-73.  
LI Zonglin, LUO Ke. Research on adaptive parameters determination in DBSCAN algorithm[J]. Computer Engineering and Applications, 2016, 52 (3): 70-73.
- [16] AHLEMEYER-STUBBE A, COLEMAN S. Data mining concept: A practical guide to data mining for business and industry[M]. New Jersey, USA: John Wiley & Sons, Ltd, 2014: 89-92.
- [17] LEE J, HAN J, LI X. Trajectory outlier detection: A partition-and-detect framework[C]//Proceedings of IEEE 24th International Conference on Data Engineering. [S. l.]: IEEE, 2008: 4497422.
- [18] 杜实.空中交通监视服务[M].北京:中国民航出版社,2012: 118-119.  
DU Shi. ATS surveillance service[M]. Beijing: China Civil Aviation Press, 2012: 118-119.

(编辑:张蓓)