

航迹点特征的时间窗分割算法的航迹聚类

王莉莉, 彭 勃[✉]

(天津市空管运行规划与安全技术重点实验室, 天津, 300300)

摘要 现有航迹聚类算法未考虑到航空器航向变化和高度下降等因素对聚类结果的影响,同时聚类过程中缺乏时间信息,另外实测二次雷达数据中存在离群点异常数据,离群点的存在会影响最终的聚类效果,使得聚类结果不准确。提出基于航迹点特征的时间窗分割算法,将航空器进场的航向变化值以及高度下降值作为确定聚类簇大小的影响因素,对进场航空器航迹点数量进行时间窗分割。对真实的进场二次雷达数据仿真分析,从仿真结果中可以看出当影响因子 a 为 0.4 时,航迹的曲率最小,聚类效果最好,进而采用层次聚类算法对不同 LOF 值所对应的航迹点进行聚类,得到最后的聚类结果可以为管制员现场指挥提供技术指导。

关键词 航迹聚类;时间窗分割算法;离群点;LOF 值

DOI 10.3969/j.issn.1009-3516.2018.03.004

中图分类号 V328.3 **文献标志码** A **文章编号** 1009-3516(2018)03-0019-05

Research on Track Clustering of Time Window Segmentation Algorithm Based on Track Point Characteristics

WANG Lili, PENG Bo[✉]

(Tianjin Key Laboratory of ATC. Operation Planning and Safety Technology, Tianjin 300300, China)

Abstract: Aimed at the problem that the existing track clustering algorithm does not take into account the influence of aircraft heading change and altitude drop on the clustering results, at the same time, the clustering process lacks time information, and in addition, outlier data exist in the two radar data, and the existence of outliers will affect the final clustering result, leading to inaccurate clustering results, this paper proposes a time window segmentation algorithm based on feature point track. The algorithm is to take the aircraft heading change approach value and height decreased value as the influence factors of determining cluster size, the paper segments the number of aircraft entering the track point time window segmentation. The simulation analysis of real approach two radar data shows that when the influence factor is 0.4, the curvature of track is minimum, and the clustering effect is the best, then the hierarchical clustering algorithm is used to cluster the tracks corresponding to different values, and the final clustering results are obtained, providing technical guidance for the controller's scene command.

Key words: track clustering; time window segmentation algorithm; outlier; LOF value

收稿日期: 2017-05-10

基金项目: 国家自然科学基金联合资助项目(U1633124)

作者简介: 王莉莉(1973—),女,陕西兴平人,教授,主要从事空中交通管理、空域规划研究. E-mail: llwang@cauc.edu.cn

通信作者: 彭 勃(1991—),男,辽宁黑山人,硕士生,主要从事航迹聚类分析研究. E-mail: 1961687423@qq.com

引用格式: 王莉莉, 彭勃. 航迹点特征的时间窗分割算法的航迹聚类[J]. 空军工程大学学报(自然科学版), 2018, 19(3): 19-23. WANG Lili, PENG Bo. Research on Track Clustering of Time Window Segmentation Algorithm Based on Track Point Characteristics [J]. Journal of Air Force Engineering University (Natural Science Edition), 2018, 19(3): 19-23.

航迹聚类分析是航迹数据分析中的核心问题,通过聚类算法对二次雷达数据中的航迹点进行聚类分析,在聚类的基础上提取出中心航迹,也即盛行交通流,将分析提取出的盛行交通流结果应用到改善终端区的进场程序管制中,可以用来改进并优化现有进离场飞行程序。国内外许多学者对航迹历史数据进行了聚类分析,例如 Gariel M 等^[1]进行了雷达航迹聚类,并对航迹运行进行实时监控;Rehm F 等^[2]直接利用已有航迹定义其相似性,针对同一机场不同跑道的进近程序聚类;Leiden K 等^[3]针对机场的繁忙区域,采用基于网格的聚类方法对不同机场的航迹进行分析,得到其不同机场的进离场中心航迹;王超^[4]等根据对实际航迹数据特征进行分析,建立起了基于对应雷达航迹点逆向比对方法的航迹相似性测度模型,同时建立了管制适用性的评价指标;王增福^[5]等对距离、方位角等平面坐标数据进行了相关处理,同时也对雷达数据提出了基于减法聚类的自适应航迹聚类算法;王涛波、黄宝军^[6]首先分析了广播式自动相关监视(ADS-B)航迹数据的特点,对异常数据进行了相关处理,并使用模糊聚类法直接提取航迹特征点。

本文将 LOF 算法、基于时间窗分割的 K 均值聚类算法以及层次聚类算法结合在一起,提出一种基于航迹点特征的时间窗分割算法,算法中将航空器的航向变化以及高度下降率作为分割时间窗内航迹点个数的影响因子,同时应用层次聚类算法对不同 LOF 值对应航迹点进行聚类,最后通过对进场的二次雷达数据进行仿真分析,验证了新方法的有效性和可行性。

1 数据初始处理

1.1 原始二次雷达数据转码

本文要处理的数据源是原始二次雷达数据。首先对原始二次雷达数据进行解码转换成全文本明码类型,然后再将所有含有有用信息的文本条目按关键字逐一检出,并将所需字段转换成程序需要的二进制整数或浮点数格式,数据单位保留不变,最后整理得到的规范航迹数据形式见表 1。

表 1 整理得到的二次雷达数据坐标

Tab. 1 Data coordinates of two radars obtained

时间	纬度/(°)	经度/(°)	高度/m	速度/ (km·h ⁻¹)	航向/(°)
11:40:26	112.289 0	32.1376	274	279	169.00
11:40:30	112.289 0	32.1376	274	226	183.00
11:40:34	112.287 0	32.1293	426	247	175.00

1.2 UTM 投影方式

由于二次雷达返回的是经纬度信息,只有转换

成大地坐标才能准确反映出飞机的实际运行轨迹,以机场导航台为坐标原点,建立直角坐标系,通过斜墨卡托投影将返回航迹的经纬度数据转换为大地坐标,采用 matlab 自带函数 UTM 进行转化。坐标转换得到直角坐标系下的横纵坐标见表 2。

表 2 经纬度坐标转换成平面坐标

Tab. 2 The longitude and latitude coordinates are converted into plane coordinates km

	1	2
1	5.545 3e+04	3.565 2e+06
2	5.545 3e+04	3.565 2e+06
3	5.533 2e+04	3.564 5e+06

2 离群点识别和剔除

2.1 LOF 算法

新算法的第一个阶段是调用 LOF 算法。假设数据集 D 是在给定度量空间内的一个有限集合, p 、 q 和 o 为 D 内的实例对象, k 为任意正整数,距离函数 $d(q, p)$ 表示对象 p 和 q 之间的距离,实例 p 的 LOF 分数通过下式来计算。

$$F_k(p) = \frac{\sum_{o \in N_k(p)} \frac{l_r d_k(o)}{l_r d_k(p)}}{|N_k(p)|} \quad (1)$$

1) 实例 p 到其第 k 个最近邻居的距离记 $k - e(p)$,若 $k - e(p) = d(p, o)$,则 p 和 o 满足:①至少有 k 个对象 $o' \in D \setminus \{p\}$,且 $d(p, o') \leq d(p, o)$ 。②至多有 $k - 1$ 个对象 $o' \in D \setminus \{p\}$,且 $d(p, o') \leq d(p, o)$ 。

2) 落在 $k - e(p)$ 范围内的数据对象,都在 p 的最近 k 个邻居集合中:

$$N_{k-e(p)} = \{q \in D \setminus \{p\} \mid d(p, q) \leq k - e(p)\} \quad (2)$$

3) 对象 p 与对象 o 之间的可达性距离定义为:
reach-dist_k(p, o) = max{ $k - e(o)$, $d(p, o)$ } (3)

4) 对象 p 的局部可达性密度计算如下:

$$d_k = 1 \left/ \left[\frac{\sum_{o \in N_k(p)} \text{reach-dist}_k(p, o)}{|K_k(p)|} \right] \right. \quad (4)$$

首先对航迹点中的离群点进行识别和剔除,计算出每个航迹点所对应的 LOF,航迹点中的 LOF 值越接近 1,说明 p 的其邻域点密度差不多, p 和邻域同属一簇;如果这个比值越大于 1,说明 p 的密度小于其邻域点密度, p 为异常点。为了给管制员提供航空器容易出现偏离的区域,在所有剔除离群点后的正常点中将 LOF 小于 0.6 的航迹点筛选出来,对其进行层次聚类,将不同类别的航迹点用不同颜色标注,并标注出航空器容易出现偏离的区域。

2.2 新算法对特征点的提取和聚类效果评价

将进场航空器的飞行看作一个过程,首先对于某一航段的航空器建立时间窗,取1 min为单位,对同一航路上的航迹点进行时间窗分割。

2.2.1 航迹点的特征提取

选取表1中的航向和高度2个指标,建立模型:每架航空器的航迹数据由一系列离散的航迹点组成,设进场时间为 T_{in} ,设某架进场航空器所对应的航迹序列为 L_k ,其中, k 为各架航空器对应的呼号, n 为 L_k 所对应的航迹点的个数。每个航迹点都包含1个6维的向量, $s_i=(t_i, x_i, y_i, z_i, v_i, h_i)$,其中 t_i 代表航迹点 s_i 对应的时刻, x_i 代表航迹点 s_i 对应的横坐标, y_i 代表航迹点 s_i 对应的纵坐标, z_i 代表航迹点 s_i 对应的飞行高度, v_i 代表航迹点 s_i 对应的速度, h_i 代表航迹点 s_i 对应的航向。对于进场航班信息中,航向变化以及高度下降率是反应航空器运行规律以及刻画航迹特征的2个重要指标,首先对航迹点中的航向变化以及高度下降率两大特征进行提取分析,建立参数:航向变化率 $\frac{\Delta h}{h}$ 、高度下降率 $\frac{\Delta z}{z}$ 以及进场航向变化率以及高度下降率的影响因子 p (以下简称影响因子),如下式,其中 $0 \leq \alpha \leq 1$:

$$p = \sqrt{\alpha \left(\frac{\Delta h}{h}\right)^2 + (1-\alpha) \left(\frac{\Delta z}{z}\right)^2} \quad (5)$$

构建权重因子矩阵 g ,对每列因子求和,得到矩阵 w :

$$g = \begin{bmatrix} p_{11} & p_{1n} \\ p_{21} & p_{2n} \\ \vdots & \vdots \\ p_{n1} & p_{nm} \end{bmatrix} \quad (6)$$

$$w = [w_1 \quad w_2 \quad \cdots \quad w_n]^T, \quad w_{i,j} = \sum_{i=1}^n p_i \sum_{j=1}^n p_j,$$

$$p_i = w_i / w \quad (7)$$

得到单位时间窗内的航迹点数:

$$n_i = \text{round}(w_i N_{LOF}) \quad (8)$$

对第 i 个时间窗 T_i 内的航迹点:

1)从时间窗内的航迹点中随机选取一个点作为初始的聚类中心, $C_i(x_{c_i}, y_{c_i}, z_{c_i})$ 。

2)计算各点与中心的距离:

$$d_{c,j} = \sqrt{(x_{c_i} - x_j)^2 + (y_{c_i} - y_j)^2 + (z_{c_i} - z_j)^2}, \quad j=1,2,\dots,n \quad (9)$$

3)不断重复计算第2)步,直到聚类中心不再发生变化或者 \bar{d}_i 逐渐收敛满足要求为止。

2.2.2 聚类效果评价

为了评价航迹聚类效果,引入曲率概念,使用最小二乘法多项式对XY方向的曲线进行拟合,通过最小化误差的平方和寻找数据的最佳函数匹配,并

使得这些求得的数据与实际数据之间误差的平方和为最小,将聚类航迹所得到的中心点数据拟合成函数表达式,再对函数求解曲率,求解出函数中曲率最小点所对应的曲率值。

3 仿真结果及分析

采用Matlab8.3环境编程进行仿真,选取某航班25条进场航迹,进场时间 $T_{in}=18$ min,见图1,从图中可以看出异常数据点有8个,剔除异常数据得到正常航迹点。利用层次聚类算法对正常航迹点进行分类,见图2,图中横坐标代表样本编号,纵坐标代表分类偏差,偏差值小于0.5的样本数过于集中,取偏差为0.8,横向对样本进行剪枝,分成3类,用3种颜色进行标注,将航迹点所对应的 α 从0.1到1.0每隔0.1取值,通过不同 α 值分割航迹点,以 $\alpha=0.1$ 时的航迹为例,对航迹进行拟合得: $f = (-0.436\ 07)x^4 + (209.764\ 7)x^3 + (-33\ 312.669\ 8)x^2 + (1\ 749\ 964.955\ 7)x^1$,求得该航迹的最小曲率为0.889 4,同样对其他进行仿真得到影响因子 α 与航迹曲率值的关系见图3,从图中可以看出当 α 取0.4时,对应的航迹曲率值最小,光滑程度最佳,进而利用影响因子 $\alpha=0.4$ 时分割时间窗得到的航迹点与进场时间窗的关系见图4, $\alpha=0.4$ 时分割时间窗得到的中心航迹见图5。

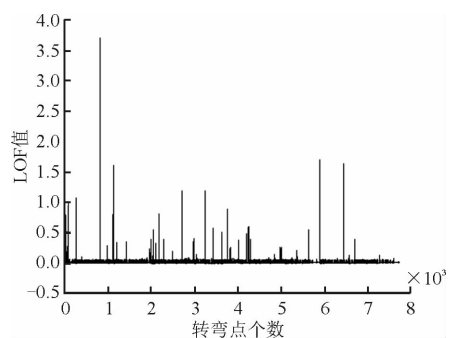


图1 离群点对应 LOF 值

Fig. 1 Outliers correspond to LOF values

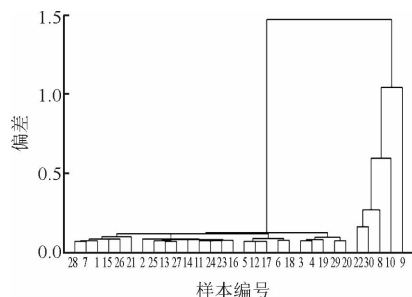


图2 层次聚类得到的 LOF 值的分类

Fig. 2 Hierarchical clustering is used to obtain the classification of the LOF values

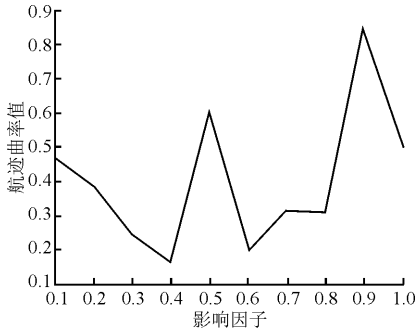


图3 影响因子与航迹曲率值的关系

Fig. 3 Relationship between influence factor and track curvature value

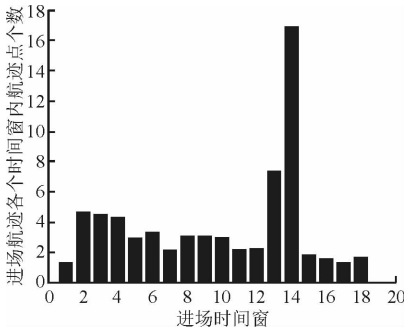


图4 $\alpha=0.4$ 时分割时间窗得到的航迹点数

Fig. 4 The segmentation time window gets the number of tracks when $\alpha=0.4$

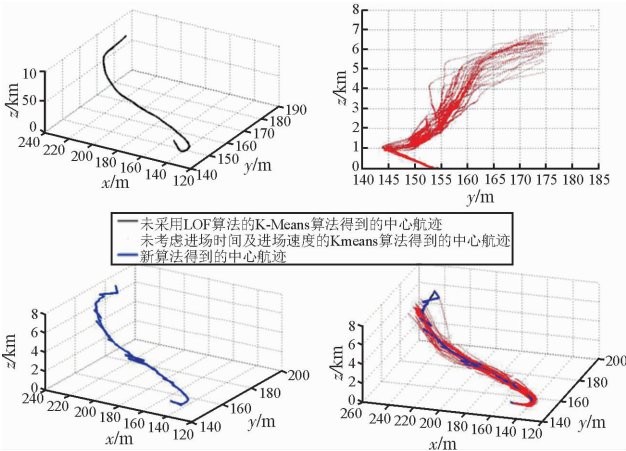


图5 聚类效果对比

Fig. 5 Clustering effect comparison

图中黑色实线代表新算法聚类得到的中心航迹,图中红色实心点代表经过 LOF 算法剔除掉离群点的航迹点,图中蓝色实线代表 K-Means 算法聚类得到的中心航迹,从图中可以看出 K-Means 算法得到的中心航迹有较多折线,不能真实地体现出航空器的进场规律,而应用新算法得到的中心航迹尤其是转弯点处的曲线更加光滑,更加符合航空器实际的运行规律。在得到图 6 中心航迹之后,由图 2 层次聚类剪枝得到 3 类 LOF 点,分别是 0~0.2、0.2~0.4、0.4~0.6,红色圆圈标注的航迹点对应的

LOF 值在 0 与 0.2 之间,将红色圆圈集中且偏离中心航迹程度最大的区域用紫色矩形框进行标注,表明该区域是进场航空器容易偏离程序的区域,为管制员现场指挥提供指导。

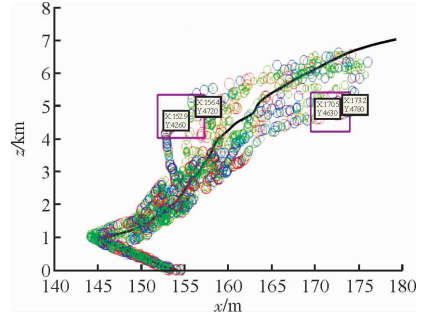


图6 容易偏离进场程序的区域标注

Fig. 6 Mark areas that are prone to deviation from entry procedures

4 结语

本文提出基于航迹点特征的时间窗分割航迹聚类方法,选取航向变化率和高度下降率作为分割航迹点时间窗的影响因子,通过真实的二次雷达进场数据仿真得到的航迹更加平滑,使用层次聚类算法对航迹点对应的 LOF 值聚类,得到容易发生位置偏移的区域,研究结果可以为管制员现场指挥提供更多地的技术指导,对进场航迹点聚类问题具有一定参考意义。

参考文献(References):

[1] GARIEL M, SRIVASTAVA A. Trajectory Clustering and An Application to Airspace Monitoring. IEEE Transactions on Intelligent Transportation Systems, 2011,12(4):1 1511-1524.

[2] REHM F. Clustering of Flight Tracks [Z]. AIAA 2010-3412.

[3] LEIDEN K, ATKIN S. Trajectory Clustering for Metroplex Operations[Z]. AIAA 2011-7066.

[4] 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):130-133.
WANG C, XU X H, WANG F. ATC Serviceability Analysis of Terminal Arrival Procedures Using Trajectory Clustering [J]. Journal of Nanjing University of Aeronautics and Astronautics, 2013, 45(1): 130-133. (in Chinese)

[5] 王增福,潘泉,郎林,等.基于减法聚类的动态航迹聚类算法[J]系统仿真学报,2009,21(16):5240-5246.
WANG Z F, PAN Q, LANG L, et al. Dynamic Track Cluster Algorithm Based on Subtractive Clustering

- [J]. Journal of System Simulation, 2009, 21(16): 5240-5 246. (in Chinese)
- [6] 王涛波,黄宝军. 基于4D航迹的模糊聚类分析[J]. 交通信息与安全, 2013, 31(180): 38-42.
WANG T B, HUANG B J. Fuzzy Cluster Analysis of 4D Trajectory Based on ADS-B [J]. Journal of Transport Information and Safe, 2013, 31(180): 38-42. (in Chinese)
- [7] 郑乐, 隋东, 张军峰, 等. 基于转弯点聚类的航空器飞行轨迹分析[J]. 武汉理工大学学报(交通科学与工程版), 2015, 39(1): 139-143.
ZHENG L, SUI D, ZHANG J F, et al. Analysis of the Aircraft Flight Path Based on Turning Points Clustering[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2015, 39(1): 139-143. (in Chinese)
- [8] 吴丹, 冯新喜. 多雷达多目标航迹起始算法研究[J]. 空军工程大学学报(自然科学版), 2006, 7(1): 16-19.
WU D, FENG X X. A Practical Algorithm of Multi-Radar Multi-Target Track Initiation [J]. Journal of Air Force Engineering University(Natural Science Edition), 2006, 7(1): 16-19. (in Chinese)
- [9] 王茜, 刘书志. 基于密度的局部离群数据挖掘方法的改进[J]. 计算机应用研究, 2014, 31(6): 1693-1696.
WANG Q, LIU S Z. Improvement of Local Outliers Mining Based on Density[J]. Application Research of Computers, 2014, 31(6): 1693-1696. (in Chinese)
- [10] 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 T, LI Y X, LYU Z P. Research on Flight Tracks Clustering Based on the Vertical Distance of Track Points [J]. Journal of Systems Engineering and Electronics, 2015, 37(9): 2198-2204. (in Chinese)
- [11] LEE J G, HAN J. Trajectory Clustering: A Partition-and-Group Framework[C]. In Proceedings 2007 ACM SIGMOD International Conference on Management of Data. Beijing, China, June 2007: 593-604.
- [12] 宫峰勋, 戴丽华, 马艳秋. 自适应选取聚类中心 K-means 航迹起始算法[J]. 哈尔滨工业大学学报, 2014, 46(5): 113-119.
GONG F X, DAI L H, MA Y Q. Algorithm of Multi-Radar Multi-Target Track Initiation Based on Adaptive K-means Clustering[J]. Journal of Harbin Institute of Technology, 2014, 46(5): 113-119. (in Chinese)
- [13] 吴笛, 杜云艳. 基于密度的轨迹时空聚类分析[J]. 地球信息科学学报, 2015, 17(10): 1162-1171.
WU D, D Y Y. Density-Based Spatiotemporal Clustering Analysis of Trajectories [J]. Journal of Geo Information Science, 2015, 17(10): 1162-1171. (in Chinese)
- [14] 赵元棣, 孙禾. 终端区航迹簇的中心航迹提取方法研究[J]. 图学学报, 2014(3): 379-386.
ZHAO Y D, SUN H. Research on Central Trajectory Extraction of Trajectory Cluster In Terminal Area [J]. Journal of Graphics, 2014(3): 379-386. (in Chinese)
- [15] 王洁宁, 孙禾, 赵元棣. 基于时间-空间的进场航迹聚类分析 [J]. 科学技术与工程, 2013(33): 10078-10081.
WANG J N, SUN H, ZHAO Y D. Approach Trajectory Clustering Analysis Based on Time-Space [J]. Science Technology and Engineering, 2013(33): 10078-10081. (in Chinese)
- [16] 徐涛, 陈雪蕊. 基于航迹聚类的终端区飞行程序轨迹表示 [J]. 四川大学学报(工程科学版), 2016(6): 188-196.
XU T, CHEN X R, et al. Flight Procedure Representation of Terminal Area Based on Flight Tracks Clustering [J]. Journal of Sichuan University (Engineering Science Edition), 2016(6): 188-196. (in Chinese)
- [17] 袁立罡, 胡明华, 张洪海, 等. 融合先验经验聚类的终端区交通流相态识别 [J]. 交通运输工程学报, 2016, 16(5): 83-94.
YUAN L G, HU M H, ZHANG H H, et al. Phase-State Identification of Traffic Flow in Terminal Area Incorporated with Prior Experience Clustering [J]. Journal of Traffic and Transportation Engineering, 2016, 16(5): 83-94. (in Chinese)
- [18] 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 C, HAN B C, WANG F. 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. (in Chinese)

(编辑: 姚树峰)