

DOI:10.16356/j.1005-2615.2020.06.009

基于聚类分析的航空器滑行过点时间预测

刘金安¹, 汤新民¹, 胡钰明², 陈强超³

(1. 南京航空航天大学民航学院, 南京, 211106; 2. 中南民航空管通信网络科技有限公司, 广州, 510080; 3. 中国民用航空中南地区空中交通管理局, 广州, 510080)

摘要: 为了预测航空器滑行预计到达时间(Estimated time of arrival, ETA), 减少场面冲突, 提高机场运行效率, 本文使用卡尔曼滤波算法对场面历史轨迹数据进行预处理。为了衡量轨迹样本间的距离, 综合三类特征用于机场场面历史轨迹数据聚类。特征包含航空器滑行时段和场面航空器数量, 以及参考动态时间规整(Dynamic time warping, DTW)算法提取的轨迹差异度特征。将两个样本特征的欧式距离作为样本间的相似度量; 基于均差最大原则确定初始聚类中心, 使用K-means算法对样本进行聚类, 根据待规划航空器的所处时段和场面航空器数量选择匹配度最高的类簇, 将其聚类中心样本的轨迹序列和塔台规划的静态路径相结合预测航空器滑行ETA。通过将实际轨迹数据与预测的滑行ETA进行对比分析, 证明了本文预测航空器滑行ETA的准确性。

关键词: 滑行预计到达时间预测; 特征提取; 历史轨迹数据; K-means 算法

中图分类号: V355

文献标志码: A

文章编号: 1005-2615(2020)06-0903-09

Prediction of Aircraft Taxiing Estimated Time of Arrival Based on Cluster Analysis

LIU Jin'an¹, TANG Xinmin¹, HU Yuming², CHEN Qiangchao³

(1. College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing, 211106, China; 2. Middle South Air Traffic Control Network Technology Co., Ltd, Guangzhou, 510080, China; 3. Air Traffic Management Bureau, Central and Southern China, Civil Aviation, Guangzhou, 510080, China)

Abstract: To predict aircraft taxiing estimated time of arrival (ETA), reduce conflicts, and improve the efficiency of the airport, this paper uses the Kalman filter algorithm to preprocess historical trajectory data of an airport. In order to measure the distance between trajectory samples, three kinds of features are synthesized for clustering of airport historical trajectory data, including the aircraft taxiing period and the number of surface aircraft. In addition, we refer to the dynamic time warp (DTW) algorithm to extract the difference features of the trajectory. The Euclidean distance of two sample features is taken as their similarity. Initial clustering center is determined based on the divided difference maximum principle. The K-means algorithm is used to cluster samples, and the cluster with the highest matching degree is selected according to the time period of the aircraft to be planned and the number of aircraft operating on the surface. The trajectory sequence of the cluster's center sample and the static path planned by the tower are combined to predict the aircraft taxiing ETA. By comparing and analyzing the actual trajectory data with the predicted taxiing ETA, the good performance of predicting the aircraft taxiing ETA in this paper is verified.

Key words: forecast taxiing estimated time of arrival (ETA); feature extraction; historical trajectory data; K-means algorithm

基金项目: 国家自然科学基金(61773202)资助项目; 四川省科技计划(2018JZ0030)资助项目。

收稿日期: 2020-07-02; **修订日期:** 2020-12-01

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: 刘金安, 汤新民, 胡钰明, 等. 基于聚类分析的航空器滑行过点时间预测[J]. 南京航空航天大学学报, 2020, 52(6): 903-911. LIU Jin'an, Tang Xinmin, Hu Yuming, et al. Prediction of aircraft taxiing estimated time of arrival based on cluster analysis[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(6): 903-911.

截至2019年,我国已有238个颁证运输机场,旅客吞吐量累计达到13.52亿人次,较2018年增长6.9%;起降架次累计达到1 166.05万架次,较2018年增长5.2%^[1]。目前对机场运转效率造成影响的主要因素包含航空器在滑行道滑行时间过长、各种滑行冲突导致的航空器多次等待以及塔台管制人员指挥飞机过多引起的各种差错等^[2]。对航空器滑行时的过点时间进行预测是实现动态路径规划的基础,为提高大型机场的整体运行效率,对航空器滑行预计到达时间(Estimated time of arrival, ETA)进行研究能够提高预测航空器到达或离开某个交叉口时间的准确性,能够在很大程度上避免航空器在滑行过程中产生冲突和造成机场拥挤,从而减少航空器滑行时间消耗^[3]。现有文献对航空器滑行ETA进行预测的方法有3类。(1)建立航空器力学模型进行预测。顾宏斌和廖超伟对处于滑行阶段的航空器建立力学模型进行六自由度分析^[4-5]。Wu等考虑到外界因素如机载操作系统对航空器滑行轨迹的影响,并提出用数学公式表达航空器滑行的动作意图,并建立动力学模型对航空器滑行轨迹进行预测^[6]。这一类预测方法比较依赖航空器的一些动力学参数,这些参数的可靠性也会对航空器滑行轨迹的预测结果造成误差影响。(2)使用无参数估计方法进行预测。Wang等基用当前统计模型的自适应滤波算法,认为加速度变化造成了滑行过程中相邻速度之间的估计误差,并对该加速度变化进行统计分析,使得研究目标自适应的调整自身加速度^[7]。宫淑丽等研究的是综合考虑无迹卡尔曼滤波和交互式多模型滤波,以进港航空器为例对其进行跟踪建模^[8]。使用无参数估计方法进行预测需要建立复杂的航空器运动模型,该模型不一定能够准确描述航空器实际的滑行状态。(3)挖掘历史轨迹数据中的信息,对研究目标未来的运动状态进行预测。张尚剑等使用多项式模型对历史数据进行处理分析,去预测研究目标接下来的位置,并且重复这个过程,实现了实时预测跟踪^[9]。该文献并未考虑历史数据的不连续问题,不一定能够准确的预测目标的位置。王莉莉等提出LOFC算法,利用基于时间窗分割的K-means算法对进场航空器航迹进行研究,期间考虑到了航空器进场平均速度,却没有考虑航空器飞行过程中速度变化趋势和终端区航空器数量对聚类结果的影响^[10]。

以上文献都为航空器滑行ETA预测提出了新颖的研究方向。基于力学模型和无参数估计方法的预测思路,通常会受到现实条件限制导致预测结

果与实际结果相差较大;而基于历史数据的预测方法没有提取足够多的特征因子参与样本相似度度量。本文首次将时段特征、场面航空器数量特征以及轨迹差异度特征进行结合,度量场面滑行历史轨迹之间的相似度。着重阐述了如何对历史轨迹数据进行特征提取以及对特征进行归一化处理统一量纲,并对经过预处理的历史轨迹数据进行聚类。根据待规划航空器的所处时段和场面航空器数量选择匹配度最高的样本,结合静态路径预测航空器滑行ETA。

1 历史数据预处理

1.1 基于卡尔曼滤波算法的去噪处理

广播式自动相关监视(Automatic dependent surveillance-broadcast, ADS-B)系统发送的报文包含的信息有航班号、时间、经度、纬度、高度、地速方向角和速度等,具有周期性和自发性^[11]。ADS-B报文经过解析后以文本格式分别存储每个航班的轨迹信息,并以对应的航班号进行命名,文本中的一行数据表示此航班的一个历史轨迹点信息。因为ADS-B数据周期约为1 s,所以航空器的轨迹由离散的轨迹点组成。本课题研究的是场面航空器滑行轨迹。先将高度不为零的轨迹点去除。由于ADS-B设备在传输数据的过程中受到天气或者设备异常的影响,会造成传输的数据异常,如轨迹数据缺失或者偏离滑行道。为消除异常数据对聚类结果的影响,需要对剩余的数据进行去噪处理。

卡尔曼滤波基本思想是在已知之前时刻的目标状态的基础上,根据目标运动的空间状态模型和观测模型。在已知之前目标状态的基础上对现在的目标状态进行估计得到一个预测值,然后再使用现在的观测信息去更新预测值,最终计算出现在的目标状态。由于本文研究的是WGS84坐标系下的轨迹坐标,所以在滤波处理前将WGS84坐标转换为以机场基准点为原点、跑道方向为X轴方向、垂直跑道方向为Y轴方向的平面坐标,同时将速度结合地速方向角和跑道方位分解为X轴和Y轴方向的分量。

本文的二维运动模型描述为

$$\mathbf{X}(k+1) = \mathbf{X}(k) + \mathbf{X}'(k)T + \frac{1}{2}\mathbf{X}''(k)T^2 \quad (1)$$

$$\mathbf{X}'(k+1) = \mathbf{X}'(k) + \mathbf{X}''(k)T \quad (2)$$

$$\mathbf{X}''(k+1) = \mathbf{X}''(k) \quad (3)$$

式中: $\mathbf{X}(k)$ 为目标在二维坐标中的位置; $\mathbf{X}'(k)$ 为目标在二维空间中的速度; $\mathbf{X}''(k)$ 为目标在二维空间中的加速度, T 为目标前后位置的时间差。目标

的状态方程为

$$X(k+1) = AX(k) + BU(k) + W(k) \quad (4)$$

式中: A 为状态转移矩阵, $X(k+1)$ 为 $k+1$ 时刻的目标状态向量, B 为输入控制矩阵, $U(k)$ 为控制信号, $W(k)$ 为均值为0、协方差为 Q 且互不相关的高斯白噪声序列。目标的观测方程为

$$Z(k+1) = HX(k+1) + V(k+1) \quad (5)$$

式中: $Z(k+1)$ 为 $k+1$ 时刻的观测向量; H 为观测矩阵, $V(k+1)$ 为协方差为 R 且互不相关的白色高斯观测噪声序列。当: $X(k) = [x(k), v_x, a_x, y(k), v_y, a_y]^T$ 时,目标的状态转移矩阵为

$$A = \begin{bmatrix} 1 & T & \frac{1}{2}T^2 & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & \frac{1}{2}T^2 \\ 0 & 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

式中: $x(k), v_x, a_x$ 分别为航空器历史轨迹点经过坐标转换后 k 时刻的横轴坐标、横向速度、横向加速度; T 为两个相邻历史轨迹点时间戳的差值。当观测向量 $Z(k+1) = [x(k), v_x, y(k), v_y]^T$ 时,观测矩阵为

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (7)$$

表示由状态向量 $X(k+1)$ 向观测向量 $Z(k+1)$ 的转换。对于每一条历史轨迹数据,采用前两个轨迹的纵横坐标和速度分量设置初始状态向量

$$X(1) = [x_1, v_{x1}, 0, y_1, v_{y1}, 0]^T \quad (8)$$

$$X(2) = \left[x_2, v_{x2}, \frac{v_{x2} - v_{x1}}{2}, y_2, v_{y2}, \frac{v_{y2} - v_{y1}}{2} \right]^T \quad (9)$$

初始滤波的误差协方差矩阵为

$$P = \begin{bmatrix} 1 & \frac{1}{T} & \frac{1}{T^2} & 0 & 0 & 0 \\ \frac{1}{T} & \frac{2}{T^2} & \frac{3}{T^3} & 0 & 0 & 0 \\ \frac{1}{T^2} & \frac{3}{T^3} & \frac{6}{T^4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{T} & \frac{1}{T^2} \\ 0 & 0 & 0 & \frac{1}{T} & \frac{1}{T^2} & \frac{3}{T^3} \\ 0 & 0 & 0 & \frac{1}{T^2} & \frac{3}{T^3} & \frac{6}{T^4} \end{bmatrix} \quad (10)$$

滤波算法从 $k=2$ 开始递推工作。对所有历史轨迹数据进行平滑处理。

1.2 特征提取

1.2.1 时段特征和场面航空器数量特征

机场不同的运行时段,航班计划数量也明显不同,提取航空器滑行时段特征有助于分析航空器的滑行规律。每个航空器在场面滑行时传回的ADS-B数据通常是半小时以内的数据,只有少数航空器传回的场面数据包含1h以上的数据。将0点至24点按小时分为24个时段,对于每个航空器的轨迹数据,判断第一个轨迹点时间戳即航空器落地时间 t 所处时段 t_s 。由于时段具有周期性,所以将其转换成周期变量,时段特征计算公式为

$$T_s = \sin\left(\frac{2\pi}{24} \cdot t_s\right) \quad (11)$$

考虑到某些时段的 T_s 相等,增加一个时段特征

$$T_c = \cos\left(\frac{2\pi}{24} \cdot t_s\right) \quad (12)$$

不同落地时间对应的时段特征如表1所示。

表1 时段特征取值示例

Table 1 Examples of feature values for periods

落地时间	时段 t_s	时段特征 T_s	时段特征 T_c
00:00~01:00	1	$\sin\left(\frac{\pi}{12}\right)$	$\cos\left(\frac{\pi}{12}\right)$
01:00~02:00	2	$\sin\left(\frac{2\pi}{12}\right)$	$\cos\left(\frac{2\pi}{12}\right)$
⋮	⋮	⋮	⋮
22:00~23:00	23	$\sin\left(\frac{23\pi}{12}\right)$	$\cos\left(\frac{23\pi}{12}\right)$
23:00~24:00	24	$\sin\left(\frac{24\pi}{12}\right)$	$\cos\left(\frac{24\pi}{12}\right)$

航空器在场面滑行时,此时场面运行航空器的数量在一定程度上能影响航空器的滑行时间,对于每个历史轨迹文本数据,获取其滑行时间区间 $[T_{start}, T_{end}]$,从当天历史数据中搜寻在 $[T_{start}, T_{end}]$ 时间范围内有轨迹点数据的航班号,统计航班数量,记为航空器数量特征 C 。

1.2.2 基于DSW算法的轨迹差异度特征

航空器在机场场面滑行的时候,以最短距离为条件规划出的静态路径,其由机场场面路径节点组成,即跑道、滑行道、停机位各边的交点。机场绘图数据库(Airport mapping database, AMDB)是描述机场拓扑关系的地理信息数据库,其中包含机场各组成元素的几何信息和特征信息。从

AMDB数据库能够获得最短路径节点的经纬度坐标 $C=[c_1, c_2, \dots, c_m]$, 其中 m 为最短路径的节点数目, $c_j=(lat_j, long_j)$ 为第 j 个节点的经纬度坐标。同样, 将其转化为以机场基准点为原点、跑道方向为 X 轴方向的平面坐标 $c_j=(a_j, b_j)$, 在历史数据中按照以下步骤匹配最短静态路径。

(1) 对于每个航班的轨迹点序列 $T=\{t_1, t_2, \dots, t_n\}$, 从 t_1 开始计算 $t_i=(x_i, y_i)(1 \leq i \leq n)$ 与 $c_1=(a_1, b_1)$ 的欧式距离 d , 定义容许误差 ϵ , 若 $d_{i1}(t_i, c_1) \leq \epsilon$, 则找到 c_1 的匹配点。

(2) 接下来从 t_{i+1} 开始继续, 按时间顺序计算与 c_2 的 d , 若 $d_{k2}(t_k, c_2) \leq \epsilon$, 其中对于 t_k , 有 $i+1 \leq k \leq n$, 则找到 c_2 的匹配点。

(3) 以此类推, 直至找到 c_m 的匹配点 $t_s(s \leq n)$ 。将从 t_i 到 t_s 所代表的轨迹点序列取出, 重复以上步骤找到所有与最短静态路径匹配的样本, 取每个样本的速度采样序列组成集合 $V=\{v^1, v^2, \dots, v^e\}$, 其中 e 是与静态最优路径匹配的样本数量。

对于速度采样序列 $v^a=[v_1^a, v_2^a, \dots, v_m^a]$ 和 $v^b=[v_1^b, v_2^b, \dots, v_n^b]$, 通常两条轨迹的轨迹点个数 m 和 n 并不相等, 不能使用欧式距离去度量 v^a 和 v^b 的相似度。动态时间规整 (Dynamic time warping, DTW) 算法通过扭曲一个时间序列, 使其每个时间点与另一个时间序列中与该时间点相差最小的时间点匹配上, 然后度量两者的相似度, 尽管两者的长度可能不同^[12]。动态空间规整 (Dynamic space warping, DSW) 算法与 DTW 算法类似, 可以在两个速度采样序列的元素顺序不发生改变的情况下, 对其中一个速度采样序列在空间轴下进行扭曲, 使得一个样本的每个元素都与另一个样本中与其距离最近的元素进行匹配^[13]。

为了对齐 v^a 和 v^b 两个速度采样序列, 需要构建一个 $m \times n$ 的矩阵网格 $A_{m \times n}$, 其中网格上面的点 (i, j) 对应的是向量中的元素 v_i^a 和 v_j^b ; 其中 $i \leq m, j \leq n$, 矩阵 $A_{m \times n}$ 中的元素 a_{ij} 表示 v_i^a 和 v_j^b 的欧式距离为

$$a_{ij} = \sqrt{(v_i^a - v_j^b)^2} \quad (13)$$

通过两条速度采样序列中相互匹配的速度元素在各自序列中的序号, 能够确定矩阵网格中对应的点, 将这些点依次连接起来即得到规整路径, 该路径记录着两条速度采样序列的匹配信息。用 $W=(w_1, w_2, \dots, w_k)$ 表示, 其中的第 k 个元素为 $w_k=(a_{ij})_k$ 。规整路径如图 1 所示, 并且规整路径

应满足如下约束

$$\begin{cases} \max\{m, n\} < K \leq m+n-1 \\ \text{边界条件: } w_1=a_{11}, w_k=a_{mn} \\ \text{连续性: } w_{k-1}=a_{i'j'}, w_k=a_{ij} \quad i-i' \leq 1; j-j' \leq 1 \\ \text{单调性: } w_{k-1}=a_{i'j'}, w_k=a_{ij} \quad i-i' \geq 0; j-j' \geq 0 \end{cases} \quad (14)$$

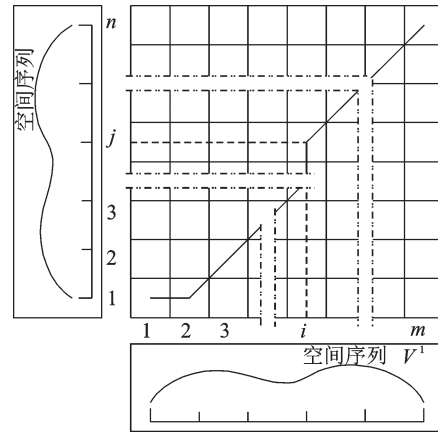


图1 规整路径示意图

Fig.1 Schematic diagram of structured path

DSW 距离是在满足以上条件的情况下, 找到两个速度采样序列 v^a 和 v^b 的元素之间累积距离的最小值, 最小距离计算公式为

$$\begin{cases} D(1, 1) = a_{11} \\ D(i, j) = a_{ij} + \min \{ D(i-1, j-1), \\ D(i, j-1), D(i-1, j) \} \end{cases} \quad (15)$$

速度采样序列 v^a 和 v^b 的相似度量

$$DSW(v^a, v^b) = \min \left\{ \sum_{i=1}^K w_i \right\} \quad (16)$$

DSW 算法步骤如下。

算法输入: 两个速度采样序列 $S[1, \dots, n]$ 和

$$T[1, \dots, m]$$

算法输出: S 和 T 的 DSW 距离 $DSW[n, m]$

For: i 的取值从 1 开始递增到 n

定义: S 中第 i 个元素和 T 中第 0 个元素的

$$DSW \text{ 距离 } DSW[i, 0] = \text{无穷大}$$

For: j 的取值从 1 开始递增到 m

定义: S 中第 0 个元素和 T 中第 j 个元素的

$$DSW \text{ 距离 } DSW[0, j] = \text{无穷大}$$

定义: S 中第 0 个元素和 T 中第 0 个元素的

$$DSW \text{ 距离 } DSW[0, 0] = 0$$

For: i 的取值从 1 开始递增到 n

For: j 的取值从 1 开始递增到 m

计算: S 第 i 个元素与 T 第 j 个元素的欧式距离

$$D(S[i], T[j])$$

S第*i*个元素与T第*j*个元素的DSW距离:

$$\begin{aligned} \text{DSW}[i,j] = & D(S[i], T[j]) + \\ & \min\{\text{DSW}[i-1,j], \text{DSW}[i,j-1], \text{DSW}[i-1,j-1]\} \end{aligned}$$

循环结束后输出DSW[n,m]

为了消除特征中量级和量纲的影响,定义轨迹差异度*D*,反映样本速度采样序列与其他样本速度采样序列的整体差异程度,对于样本λ,计算其速度采样序列 v^λ 与其余所有样本速度采样序列的DSW距离,求得其最值DSW_{MIN}、DSW_{MAX}和均值 $\overline{\text{DSW}}$,样本λ的轨迹差异度

$$D^\lambda = \frac{\overline{\text{DSW}} - \text{DSW}_{\text{MIN}}}{\text{DSW}_{\text{MAX}} - \text{DSW}_{\text{MIN}}} \quad (17)$$

类似的求出所有样本的轨迹差异度。为了消除特征之间量纲的影响,采用极差变换法将 T_s 、 T_c 、 C 归一化,特征量化公式为

$$x = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (18)$$

至此得到样本集合 $U = \{u^1, u^2, \dots, u^e\}$, $u^e = [T_s^e, T_c^e, C^e, D^e, T_r^e]^T$,其中*e*是与静态最优路径匹配的样本数量, T_s^e 和 T_c^e 为时段特征, C^e 为场面航空器数量特征, D^e 为轨迹差异度特征, T_r^e 为轨迹序列。

2 基于K-means算法的航空器滑行ETA预测

2.1 基于K-means算法的历史轨迹聚类

目前基于历史轨迹数据对航空器轨迹进行预测的方法提出的比较多。王莉莉等从航空器进场时间入手,对不同时间窗内的航迹数据进行聚类分析^[10]。李迎迎等考虑了场面运动航空器的数量,根据历史数据建立了离场滑行时间的预测模型^[14]。本文综合考虑了历史数据中航空器的降落时间和场面运动航空器数量,结合轨迹差异度特征进行聚类分析,并与单独考虑航空器降落时间或场面运动航空器数量的聚类结果进行对比。本文将采用K-means聚类算法对历史轨迹样本进行聚类,K-means聚类算法理论严密,实现简单能对数据集进行高效聚类,但如何确定聚类类数*K*值以及初始聚类中心,对于聚类结果有着很大的影响^[15-18]。样本

$u^\alpha = [T_s^\alpha, T_c^\alpha, C^\alpha, D^\alpha, T_r^\alpha]^T$ 与 $u^\beta = [T_s^\beta, T_c^\beta, C^\beta, D^\beta, T_r^\beta]^T$ 的相似度量

$$\text{DIS}_{\alpha\beta} = [(T_s^\alpha - T_s^\beta)^2 + (T_c^\alpha - T_c^\beta)^2 + (C^\alpha - C^\beta)^2 + (D^\alpha - D^\beta)^2]^{\frac{1}{2}} \quad (19)$$

若在轨迹聚类时仅考虑时段特征和轨迹差异度特征,样本 u^α 与 u^β 的相似度量

$$\text{DIS}_{\alpha\beta}^{TD} = [(T_s^\alpha - T_s^\beta)^2 + (T_c^\alpha - T_c^\beta)^2 + (D^\alpha - D^\beta)^2]^{\frac{1}{2}} \quad (20)$$

若在轨迹聚类时仅考虑场面航空器数量特征和轨迹差异度特征,样本 u^α 与 u^β 的相似度量

$$\text{DIS}_{\alpha\beta}^{CD} = \sqrt{(C^\alpha - C^\beta)^2 + (D^\alpha - D^\beta)^2} \quad (21)$$

本文采用肘部法则确定最优*K*值。肘部法通过计算每个聚类的中心样本与该聚类中每个样本的相似度之和得到该聚类的畸变程度,值越低说明聚类的样本之间越紧密,越高说明聚类的样本之间越松散^[19-20]。求出所有聚类的畸变程度并计算平均值,该值与*K*值一般呈负相关。如果样本之间有一定的区分度,该值会在某个*K*值较*K*-1时显著减小,这时的*K*值即为聚类效果相对最优的聚类类数。在此基础上,本文设置*K*值从最小值不断递增,计算相应聚类结果的平均畸变程度,找到聚类性能最优的*K*值。

定义样本*i*的均差*d_i*和样本整体均差*M*

$$d_i = \frac{\sum_{j=1}^N \text{DIS}_{ij}}{N} \quad (22)$$

$$M = \frac{\sum_{i=1}^N d_i}{N} \quad (23)$$

初始聚类中心的确定方法如下:

(1)计算每两个样本的距离,然后确定每个样本的均差和样本集整体均差*M*,将均差最大的样本作为第一个聚类中心,并将该样本从样本集中删除;

(2)继续找到当前样本集中均差最大的样本,将其从样本集中删除,如果该样本与之前确定的聚类中心样本的距离都大于*M*,将该样本作为第二个聚类中心;

(3)以此类推,直至找到*K*个聚类中心。

若需要更新聚类中心,对每个聚类中的每个样本重新计算均差

$$d_i' = \frac{\sum_{j=1}^n \text{DIS}_{ij}}{n} \quad (24)$$

式中:*n*为样本所在聚类中的样本数。接下来从每个聚类中找到均差最小的那个样本作为新的聚类中心,继续迭代,直至聚类中心不再变化时,聚类结束。

聚类性能的评价主要是考虑同簇样本与不同

簇样本的相似度。为了保证同簇样本之间的相似度尽可能高,不同簇样本之间的相似度尽可能低,有以下两种度量标准:

(1)紧凑度:用来衡量同簇样本点之间相似程度的指标。

(2)分离度:指不同簇的样本之间差异是否足够大。

本文采用轮廓系数度量聚类效果,轮廓系数能够综合上述两个度量标准评价聚类的效果,其取值范围为 $[-1,1]$ 。轮廓系数越大说明聚类的效果越好,计算公式为

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (25)$$

式中: a_i 为样本*i*与它所属簇中其他样本的平均距离; b_i 为样本*i*与它相邻最近的簇中所有样本的平均距离。求得所有样本的轮廓系数,其平均值就是聚类结果总的轮廓系数。

2.2 航空器滑行ETA生成

聚类结束后,根据航空器的待规划时间和当前场面航空器数量结合1.2节中 T_s 、 T_c 、 C 的最值,更新归一化后的待规划航空器特征 C^0 、 T_s^0 、 T_c^0 。计算与聚类结果中每个聚类中心的特征匹配距离为

$$DIS_j = \sqrt{(C - C^0)^2 + (T_s - T_s^0)^2 + (T_c - T_c^0)^2} \quad (26)$$

取最小距离对应的聚类中心,根据其轨迹序列 T_r 和静态路径生成待规划航空器的滑行ETA,静态路径中每个节点ETA相对于起点出发时间的差值等于 T_r 中对应匹配点的过点时间相对于 t_{r1} 的过点时间的差值,具体过程如下:

(1)静态路径根据*n*个路径节点可以分成*n*-1个路段,由1.2.2节知静态路径的第一个节点Node₁的匹配点 t_{r1} 为聚类中心样本的第一个轨迹点,静态路径最后一个节点Node_n的匹配点 t_{rn} 为最后一个轨迹点,由匹配点的速度也可得到每个节点的过点速度。

(2)对于待规划航空器,规划时间为起始时间,即路段1的到达时刻 t_{1start} 。类似1.2.2节中的路径匹配方法,从匹配样本中寻找Node₂的匹配轨迹点 t_{r2} , t_{r2} 的时间减去 t_{r1} 的时间即为路段1的预计通过时间 t_1 ,则路段1的离开时刻 $t_{1end} = t_{1start} + t_1$,如图2所示。

(3)路段2的到达时刻 $t_{2start} = t_{1end}$,接下来从匹配样本中寻找Node₃的匹配轨迹点 t_{r3} , t_{r3} 的时间减去 t_{r2} 的时间即为路段2的预计通过时间 t_2 ,则路段2的离开时刻 $t_{2end} = t_{2start} + t_2$ 。

(4)以此类推,可以预测每个节点的ETA,并得到待规划航空器的预测速度剖面。

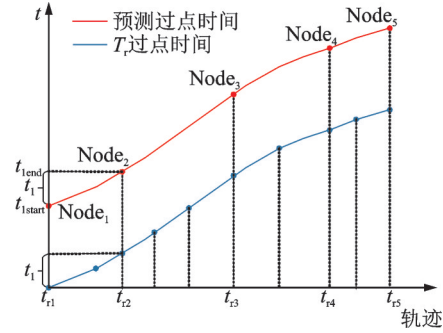


图2 结合轨迹序列和静态路径预测示意图

Fig.2 Schematic of combination of trajectory sequence and static path prediction

3 仿真实验

本文实验是以南京禄口国际机场为例,对机场2019年3月的历史轨迹数据进行预处理以及聚类,其中一条历史轨迹如图3所示。

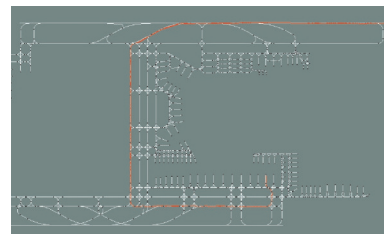


图3 原始轨迹数据示意图

Fig.3 Schematic diagram of original trajectory data

对历史轨迹数据进行去噪处理,并提取时段特征和场面航空器数量特征,特征 T_s 、 T_c 、 C 的最值如表2所示,对于图4所示的静态路径,设置容许误差 $\epsilon = 20$,根据其节点坐标在已处理的历史数据中进行匹配,计算并提取每条匹配轨迹的轨迹差异度特征,得到输入样本集合。

表2 禄口机场3月历史轨迹特征最值

Table 2 Mini and max feature values of historical trajectory in Nanjing Lukou International Airport in March

T_{sMAX}	T_{sMIN}	T_{cMAX}	T_{cMIN}	C_{MAX}	C_{MIN}
0.866	-1	1	-1	18	1

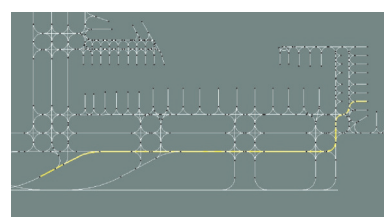


图4 静态规划路径示意图

Fig.4 Schematic diagram of paths by static planning

选取不同的 K 值,计算平均畸变程度和平均轮廓系数,由图 5,6 知在 $K=4$ 时,畸变程度得到大幅度改善,并且聚类效果相对最优。 $K=4$ 时聚类结果如表 3 所示。

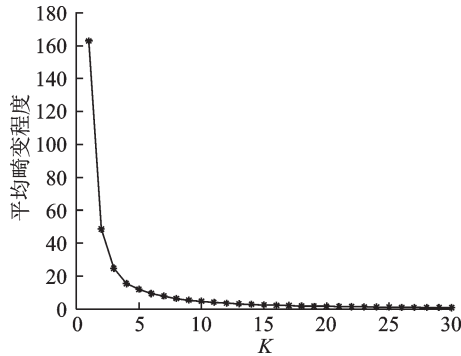


图 5 肘部法则-畸变趋势图

Fig.5 Elbow method-distortion trend

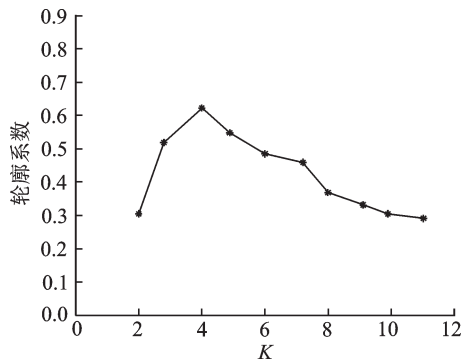


图 6 轮廓系数指标

Fig.6 Silhouette coefficient index

表 3 聚类结果

Table 3 Clustering results

聚类	C	T_s	T_c	DSW	聚类中心 耗时/s	样本 数量
1	0.471	0.018	0.371	0.097	252.423	65
2	0.706	0.072	0.250	0.100	270.447	37
3	0.529	0.804	0.067	0.119	248.540	114
4	0.412	0.157	0.854	0.136	271.800	81

取禄口机场 4 月份的 7 天数据进行验证,首先对 7 天的历史轨迹数据进行预处理,提取时段特征和场面航空器数量特征并根据表 2 的最值进行归一化处理。与图 4 所示的静态路径进行匹配后得到 76 条轨迹数据。以航班 CCA4373 的轨迹数据 $u=[T_s^0, T_c^0, C^0, T_r]^T$ 为例,其特征值为 $C^0=0.7647$ 、 $T_s^0=0.509$ 、 $T_c^0=0.9999$ 。计算与表 3 每个聚类中心的特征匹配距离 DIS_f ,结果如表 4 所示。

由表 4 中最小 DIS_f 值匹配到聚类 4,其聚类中心代表的样本所耗时间为 271.8 s。按照 2.2 节的方法计算得到 CCA4373 的预测滑行 ETA,如表 5 所示。

表 4 特征匹配距离

Table 4 Feature matching distances

聚类	1	2	3	4
DIS_f	0.850	0.870	1.006	0.519

表 5 滑行 ETA 预测结果

Table 5 Predicted results of ETA

节点	ETA/s	预测过点速度/($m \cdot s^{-1}$)
1	0.000	6.392
2	8.393	5.738
3	20.575	5.570
4	47.640	5.654
5	83.230	8.738
⋮	⋮	⋮
18	256.955	5.140
19	258.785	5.140
20	260.835	4.112
21	264.249	4.112
22	268.399	4.112

接下来采用 $DIS_{\alpha\beta}^{TD}$ 或 $DIS_{\alpha\beta}^{CD}$ 进行样本相似度度量,与采用 $DIS_{\alpha\beta}$ 进行聚类预测的过程相同,预测静态路径每个节点的 ETA 和过点速度。航班 CCA4373 的轨迹序列 T_r 经过静态路径每个节点的时间与本文预测的 3 种 ETA 对比结果如表 6 所示。为方便比较,过点时间都处理为相对起点时间的差值,其速度剖面与本文的预测速度剖面对比结果如图 7 所示。

从图 7 可以看出实际轨迹与预测轨迹的升降趋势大致相同,为了验证不同预测结果的准确性,定义平均过点时差 \bar{t} 和平均过点偏移 \bar{s}

$$\bar{t} = \frac{\sum_{i=1}^n |t_i^p - t_i^r|}{n} \quad (27)$$

$$\bar{s} = \bar{t} \cdot \bar{v} \quad (28)$$

式中: t_i^p 和 t_i^r 分别表示预测过点时间和航班 CCA4373 对应过点时间的差值; n 为静态路径节点数,由航班 CCA4373 的轨迹序列 T_r 得平均过点速度 $\bar{v} = 5.864 \text{ m/s}$,3 种预测结果的 \bar{t} 和 \bar{s} 计算结果对比如表 7 所示。

此为 76 条轨迹数据中一条轨迹的预测结果对比。类似地对剩余 75 条轨迹数据进行处理,最后对每个相似度量对应的 76 个 \bar{s} 求平均值,得到 3 种相似度量的总体平均过点偏移 \bar{S} ,结果如表 8 所示。

从表 8 中可以看出,综合考虑时段特征和场面航空器数量特征,相较于单独考虑其中一者与轨迹差异度特征结合的相似度量,能够提高聚类性能,同时簇内相似度更高,预测结果更加稳定,以禄口机场 A320 机型为例,机身长度为 37.57 m,本文采

表6 过点时间与速度计算

Table 6 Calculation of arrival time and speeds

节点	DIS _{aβ} 预测 ETA/s	DIS _{aβ} 预测速 度/(m·s ⁻¹)	DIS _{aβ} ^{TD} 预测 ETA/s	DIS _{aβ} ^{TD} 预测速 度/(m·s ⁻¹)	DIS _{aβ} ^{CD} 预测 ETA/s	DIS _{aβ} ^{CD} 预测速 度/(m·s ⁻¹)	实际过点时 间/s	实际过点速 度/(m·s ⁻¹)
1	0.000	6.392	0.000	8.878	0.000	9.850	0.000	5.140
2	8.393	5.738	8.185	8.766	7.986	9.364	11.984	5.654
3	20.575	5.570	20.285	6.168	20.173	9.392	19.586	5.654
4	47.640	5.654	47.050	7.196	46.288	9.392	35.781	6.682
5	83.230	8.738	83.180	7.196	82.165	8.822	69.397	8.224
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
18	256.955	5.14	254.934	4.626	251.819	5.710	251.933	4.626
19	258.785	5.14	256.282	4.626	253.126	5.140	256.259	4.626
20	260.835	4.112	258.815	5.140	254.455	5.654	260.852	4.112
21	264.249	4.112	262.186	5.140	258.964	5.654	266.377	4.626
22	268.399	4.112	265.276	5.140	263.582	5.654	268.224	4.112

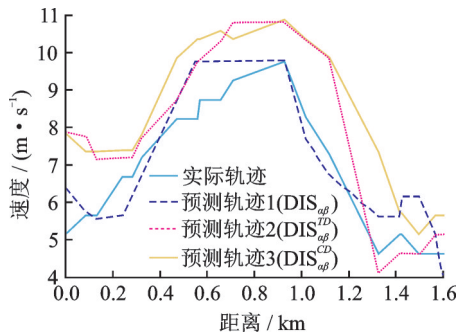


图7 预测速度剖面与实际速度剖面

Fig.7 Predicted velocity profile and actual velocity profile

表7 3种预测结果准确性对比结果1

Table 7 Comparison 1 of the accuracies of the three prediction results

预测结果	DIS _{aβ} 预测 结果	DIS _{aβ} ^{TD} 预测 结果	DIS _{aβ} ^{CD} 预测 结果
\bar{t}/s	3.983	4.621	5.342
\bar{s}/m	23.356	27.098	31.325

表8 3种预测结果准确性对比结果2

Table 8 Comparison 2 of the accuracies of the three prediction results

预测结果	DIS _{aβ} 预 测结果	DIS _{aβ} ^{TD} 预 测结果	DIS _{aβ} ^{CD} 预 测结果
总体平均过点偏移 \bar{S}	22.763	27.794	29.658
标准差 σ	1.652	2.162	2.837

用DIS_{aβ}进行预测的总体平均过点偏移约60.59%个机身长度,说明本文预测滑行ETA的准确性很高。

4 结 论

机场运行过程中产生了大量的轨迹数据,其中蕴含着许多对机场管理人员有益的信息以待挖掘。目前通过历史轨迹数据对航空器4D航迹的预

测研究比较多,而结合场面历史轨迹数据对航空器场面滑行ETA的预测研究还比较少。本文着重研究了轨迹聚类的技术。在轨迹特征提取方面综合了滑行时段、场面航空器运行数量和轨迹整体差异程度,对历史轨迹数据进行预处理和聚类分析。根据待规划航空器的规划时段和当前场面航空器数量,选择匹配度最高的类簇,结合其聚类中心代表的样本和静态路径对待规划航空器的滑行ETA进行预测。并且通过实际轨迹数据与预测的滑行ETA进行对比分析,证明了本文预测的准确性。同时相比较于现阶段基于历史数据进行轨迹预测的研究思路,将历史数据的时段特征和场面航空器数量特征用于轨迹聚类。

为了度量历史轨迹速度采样序列的差异度,本文在现有轨迹预测研究的基础上,提出轨迹差异度特征,并将其与时段特征和场面航空器数量特征结合进行滑行轨迹预测。与仅结合时段特征和场面航空器数量特征之一进行轨迹预测的结果进行比较分析,证明了本文在轨迹预测方面提出的预测模型较已有研究思路有了一定的性能提升。下一步将综合考虑场面多航空器运行状况,进行各航空器滑行ETA预测,为机场场面冲突预测和动态路由规划奠定基础。

参考文献:

[1] 闵梓. 民航局公布2019年民航行业发展统计公报[N]. 中国航空报, 2020-06-09(007).
MIN Zi. Civil Aviation administration announces 2019 statistical bulletin of civil aviation industry development[N]. China Aviation News, 2020-06-09(007).
[2] 华昕. 南京禄口国际机场航班正常性研究[J]. 江苏航空, 2017(2): 31-33.
HUA Xin. Study on the flight regularity of Nanjing Lukou International Airport[J]. Jiangsu Airlines, 2017

- (2):31-33.
- [3] 张明,韩松臣,杭勇,等.基于改进的可拓评价模型的安全管制员工作负荷综合评估模型[J].安全与环境学报,2014,14(1):84-87.
ZHANG Ming, HAN Songchen, HANG Yong, et al. Comprehensive evaluation model of safety controller workload based on improved extension evaluation model[J]. Journal of Safety and Environment, 2014, 14(1):84-87.
- [4] 顾宏斌.飞机地面运行的动力学模型[J].航空学报,2001,22(2):163-167.
GU Hongbin. Dynamic model of aircraft ground operation[J]. Journal of Aeronautics, 2001, 22(2):163-167.
- [5] 廖超伟.航空器跑道滑行轨迹预测方法研究[J].舰船电子工程,2015,35(8):127-129,166.
LIAO Chaowei. Research on the prediction method of taxiway trajectory of aircraft runway[J]. Ship Electronic Engineering, 2015, 35(8):127-129,166.
- [6] WU D, ZHAO Y J, CAPOZZI B. Fundamental surface trajectory models for air traffic automation[C]//Proceedings of the 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference. Fort Worth, USA: AIAA, 2010:1-34.
- [7] WANG W, HOU H L. An improved current statistical model for maneuvering target tracking[C]//Proceedings of the 4th IEEE Conference on Industrial Electronics and Applications. [S.l.]: IEEE, 2009:4017-4020.
- [8] 宫淑丽,王帮峰,吴红兰,等.基于IMM算法的机场场面运动目标跟踪[J].系统工程与电子技术,2011,33(10):2322-2326.
GONG Shuli, WANG Bangfeng, WU Honglan, et al. Airport scene moving target tracking based on IMM algorithm[J]. Systems Engineering and Electronic Technology, 2011, 33(10):2322-2326.
- [9] 张尚剑,刘永智.用滑动窗多项式拟合法实时预测运动目标轨迹[J].光电工程,2003(4):24-27.
ZHANG Shangjian, LIU Yongzhi. Real-time prediction of moving target trajectory using sliding window polynomial fitting method[J]. Optoelectronic Engineering, 2003(4):24-27.
- [10] 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 and Astronautics, 2018, 50(5):661-665.
- [11] 刘燕.广播式自动相关监视(ADS-B)数据处理技术研究[C]//2019第七届中国指挥控制大会论文集.北京,中国:中国指挥与控制学会,2019:272-276.
LIU Yan. ADS-B data processing technology research [C]//Proceedings of the 7th China Command and Control Conference 2019. Beijing, China: Chinese Command and Control Society, 2019:272-276.
- [12] HUANG Shih-Feng, LU Hong-Ping. Classification of temporal data using dynamic time warping and compressed learning[J]. Biomedical Signal Processing and Control, 2020, 57:101781.
- [13] 周龙,汤新民,汤森.基于速度剖面拟合的航空器场面滑行ETA预测[J].航空计算技术,2015,45(1):35-39,44.
ZHOU Long, TANG Xinmin, TANG Miao. 4D trajectory prediction of aircraft surface taxiing based on velocity profile fitting[J]. Aeronautical Computing Technology, 2015, 45(1):35-39,44.
- [14] 李迎迎,马一帆,吕成宇,等.航空器滑行时间预测分析[J].江苏航空,2020(3):11-14.
LI Yingying, MA Yifan, LU Chengyu, et al. Forecast and analysis of aircraft taxi time[J]. Jiangsu Aviation, 2020(3):11-14.
- [15] 曾俊.基于划分的数据挖掘K-means聚类算法分析[J].现代电子技术,2020,43(3):14-17.
ZENG Jun. Analysis of K-means clustering algorithm for data mining based on partition[J]. Modern Electronic Technology, 2020, 43(3):14-17.
- [16] XIA Dongliang, NING Feifei, HE Weina. Research on parallel adaptive canopy-k-means clustering algorithm for big data mining based on cloud platform[J]. Journal of Grid Computing, 2020, 18(2):263-273.
- [17] 黄吉.一种K-means聚类改进算法研究及应用[D].武汉:湖北工业大学,2018.
HUANG Ji. Research and application of an improved K-means clustering algorithm[D]. Wuhan: Hubei University of Technology, 2018.
- [18] ZHANG Shaonan, LI Shanshan, HU Jiaqiao, et al. An iterative algorithm for optimal variable weighting in K-means clustering[J]. Taylor & Francis, 2019, 48(3/4/5):1346-1365.
- [19] CHOWDHURY K, CHAUDHURI D, PAL A K, et al. Seed selection algorithm through K-means on optimal number of clusters[J]. Multimedia Tools and Applications, 2019, 78(13):18617-18651.
- [20] 郑英鑫.数据挖掘中基于肘部法则的聚类分析在中小学生学习路线优化设计的应用[J].电子世界,2017(9):146.
ZHENG Yingxin. Application of cluster analysis based on elbow rule in data mining to optimize the design of elementary and middle school students' travel routes [J]. Electronic World, 2017(9):146.