

基于 ST DBSCAN 的航迹聚类实现

朱瑜亮

(中国电科集团第十研究所,四川 成都 610036)

摘要: 针对现有飞行目标的航迹聚类算法的不足,提出的一种可对任意形状的航迹聚类,且不需提前划分聚类目标个数,可解决时空域航迹数据的聚类方法。通过航迹信息中的空间坐标以及时间信息,扫描选取航迹数据中的任意未标记点,进行时间域上的邻近点扫描,再对时间邻域内的点进行空间域扫描,通过时空域内邻近点迹的数量将相同目标的航迹形成簇,并通过数据仿真验证本聚类方法的有效性。

关键词: 空中目标;航迹聚类;ST DBSCAN;仿真;聚类参数

中图分类号: TN95

文献标识码: A

DOI:10.16157/j.issn.0258-7998.212331

中文引用格式: 朱瑜亮. 基于 ST DBSCAN 的航迹聚类实现[J]. 电子技术应用, 2022, 48(5): 125-128.

英文引用格式: Zhu Yuliang. Implementation of track clustering based on ST DBSCAN[J]. Application of Electronic Technique, 2022, 48(5): 125-128.

Implementation of track clustering based on ST DBSCAN

Zhu Yuliang

(The 10th Research Institute of CETC, Chengdu 610036, China)

Abstract: Aiming at the shortcomings of existing track clustering algorithms for flying targets, a clustering method for track data in space-time domain is proposed, which can cluster tracks of arbitrary shape without dividing the number of clustering targets in advance. Through the spatial coordinates and time information in the track information, it scans and selects any unmarked points in the track data, scan the proximity points in the time domain, and then scans the points in the time neighborhood in the space domain. Cluster the tracks of the same target through the number of proximity points in the space-time domain, and verify the effectiveness of this clustering method through data simulation.

Key words: air target; track clustering; ST DBSCAN; simulation; clustering parameters

0 引言

航迹聚类是航迹数据分析中的一个重点,通过对记录或实时的航迹数据的聚类分析,可以获得不同空中目标的飞行路径、飞行范围、飞行特征等信息,帮助指挥人员了解空中目标态势,对空情进行指挥或事后研究分析。

以往传统航迹聚类方法,有可以对空间中 K 个点为中心进行聚类,对最靠近中心点的对象归类,并迭代中心点位置的 K -means 算法。 K -means 算法可对圆形或球状的聚类数据进行处理,但无法处理任意形状的数据簇聚类,且需要事先人工设置聚类类别数 K ,聚类结果也受起始中心点设置的影响。

也有基于网格的移动对象处理算法,将时空域划分为网格,把航迹数据点划分到不同网格内可解决航迹点本身的定位误差,再将邻域密度高于门限的网格连接成簇进行聚类。但对密度不均匀、密度差异大的数据集聚类效果不好,门限参数不好选取,且不能处理空域数据的聚类。

还有基于空间密度的聚类算法 DBSCAN,通过统计点迹数据集内任意点邻域内邻近点的数量,不断向邻域扩张聚合为簇,直到遍历所有点完成聚类。同样可发现任意形状的数据聚类,但对密度不均匀、密度差异大的数据集聚类效果不好,且不能处理时空域数据的聚类。

本文使用基于 ST DBSCAN 的航迹聚类,既符合 DBSCAN 聚类算法的特点和优点又增加了从时间域对目标数据集的扫描。

1 ST DBSCAN 算法

ST DBSCAN 算法的航迹提取实现是在时间域和空间域上对目标点迹进行扫描,以目标航迹点的密度为依据进行聚类。通过此种聚类方式可对任意形状的飞行航迹进行聚类。

1.1 DBSCAN 算法

(1)把所有数据集 D 的数据点都标记为无分类标签点;

(2)从任意无标签点开始选取某点 p 为扫描的核心

点,以核心点 p 为圆心,距离邻域 Eps 为半径的圆区域 $Eps(p)$ 为扫描邻域,扫描满足邻近距离门限的邻域点数量;

(3)如果该核心点的邻域点数量小于设定阈值 $MinPts$,则标记该点为噪声点;

(4)如果邻域点数量大于等于设定阈值 $MinPts$,则标记该点 p 为核心点,产生聚类编号为簇 $C1$,并将 p 在邻域 $Eps(p)$ 内的点都加入待扫描的簇 $C1$ 中;

(5)对簇 $C1$ 内的尚未被标记的点重复步骤(2)~步骤(4)的过程,直到簇 $C1$ 内的点全部被标记处理;

(6)继续从数据集 D 内选取其他未标记过的点,重复步骤(2)~步骤(5)的过程,聚类为另一个新类簇 $C2$ 、 $C3$ 等,直到 D 内没有未标记过的点。

1.2 ST DBSCAN 算法的改进

ST DBSCAN 算法的改进在于将 DBSCAN 算法中的第(2)步加入了对扫描邻域的时间连续性约束,从而可以从时间域维度提高飞机航迹聚类的准确性。具体方法如下:

(1)把所有飞行航迹数据集 D 的数据点都标记为无分类标签点;

(2)从任意无标签点开始选取某点 p 为扫描的核心点,核心点 p 时间邻域 ΔT 内为扫描邻域,扫描时间邻域 ΔT 内满足以核心点 p 为圆心,距离邻域 Eps 为半径的圆球区域 $Eps(p)$ 中满足距离门限的邻域点数量;

(3)如果该核心点的邻域点数量小于设定阈值 $MinPts$,则标记该点为噪声点;

(4)如果邻域点数量大于等于设定阈值 $MinPts$,则标记该点 p 为核心点,产生聚类编号为簇 $C1$,并将 p 在邻域 $Eps(p)$ 内的点都加入待扫描的簇 $C1$ 中;

(5)对簇 $C1$ 内的尚未被标记的点,重复步骤(2)~步骤(4)的过程,直到簇 $C1$ 内的点全部被标记处理;

(6)继续从数据集 D 内选取未标记过的点,重复步骤(2)~步骤(5)的过程,聚类为另一个类,直到 D 内没有未标记过的点。

1.3 参数选取

1.3.1 Eps 值的选取

Eps 的值可通过计算排序 K 最近邻距离值获得。

一般情况下,数据集中的噪音点与可聚类数据应该有较大的密度差异才能对不同点迹区分聚类,在 K 最近邻距离的排序中 K 最近邻距离的值越小说明点迹密度越大,反之说明点迹密度越小。在本文中数据集中,当 K 最近邻距离值出现由小到大的明显突变点时,说明点迹的密度从较大的有效航迹区域进入到了密度较小的噪音区域,表明存在一个阈值点可以将数据集的点区分开来。具体的计算方法如下:

(1) K 最近邻距离方法是指对数据集 $D=\{p(i); i=0, 1, \dots, n\}$ 内的点,计算出点 $p(i)$ 到数据集 D 内其他每个

点的距离;

(2)对距离由小到大排序后得到集合 $Distance=\{d(1), d(2), \dots, d(k), \dots, d(n)\}$,其中顺次为 k 的距离 $d(k)$ 就是 $p(i)$ 的 K 最近邻距离;

(3)计算 D 中所有点的 K 最近邻距离的集合 $E=\{e(1), e(2), \dots, e(n)\}$,对 E 中所有距离求均值 e' ,并计算出各种 K 值下 e' 变化的值,观察变化值最急剧的部分,对应的 e' 值即为 Eps 半径。

1.3.2 MinPts 的取值

$MinPts$ 用来控制聚类簇中点迹的最小数目,可根据聚类簇中点迹的最小可能数进行设置,保证一定有点迹可以完成聚类。

1.3.3 ΔT 的取值

ΔT 用来控制聚类簇中点迹的最小邻近时间,可根据点迹数据中的最小时间间隔进行设置,保证一定有点迹可以完成聚类。

2 仿真实现

2.1 生成航迹数据

(1)在 STK 9 中建立场景产生 3 条飞机的航迹;

(2)3 条航迹数据分别导出后使用 Python 对数据预处理;

(3)提取 3 个航迹文件中的时间以及经纬高数据、添加列名到 dataframe 中;

(4)在航迹文件中添加随机噪音数据;

(5)为不同航迹、噪声数据预先打上标签,用于聚类后的结果评估。

组合以上 4 组数据并保存为测试数据集。

2.2 航迹数据特征

数据中包含了航迹数据和设置的噪声数据信息(其中经纬高信息使用 WGS-84 直角坐标系数值),并为不同航迹、噪声数据预先打上了标签:航迹 1、航迹 2、航迹 3、噪声,用于聚类后的结果评估,具体如表 1 所示。

表 1 航迹数据内容

标签	起止点时间	点迹时间间隔	点数	关系
航迹 1	0 s~285.24 s	20 ms	14 000	与航迹 2、3 独立不交叉
航迹 2	0 s~276.062 s	20 ms	13 804	与航迹 3 相交叉
航迹 3	500 s~754.70 s	20 ms	12 737	与航迹 2 相交叉
噪声点	0~800 s	随机分布	1 500	随机分布

(1)航迹 1 与航迹 2 在时域的起止点、时间间隔相同,空间域路径无交集;

(2)航迹 2 与航迹 3 在时间域的起止点不同、时间间隔相同,空间域路径有交集;

(3)航迹 1 与航迹 3 在时间域的起止点不同、时间间隔相同,空间域路径无交集。

2.3 聚类参数选取

(1) Eps 的取值

根据 K 最近邻距离方法,对数据集中航迹点迹的 $K=2\sim 24$ 取值时的 K 最近邻距离进行顺序排序,可见在 K 值在 $3\rightarrow 4$ 、 $5\rightarrow 6$ 、 $7\rightarrow 8$ 处的欧式距离值的跳变值都较大,因此以 $K=3$ 、 5 、 7 时的取值作为 Eps 值,分别为 5.522、11.251、17.191,具体如表 2 所示。

表 2 K 最近邻距离

K 值	2	3	4	5	6	7	8	...	23	24
距离值	5.303	5.522	10.818	11.251	16.544	17.191	22.481	...	72.342	77.593

(2)MinPts 的取值

根据聚类簇中点迹的最小可能数设置为 2。

(3) ΔT 的取值

根据点迹数据中的时间间隔设置为能包含 2 个点迹的时间值 0.04 s。

2.4 聚类结果

2.4.1 原始航迹数据

原始的航迹测试数据的三维图中所有的点迹未聚类前的视图,如图 1 所示。

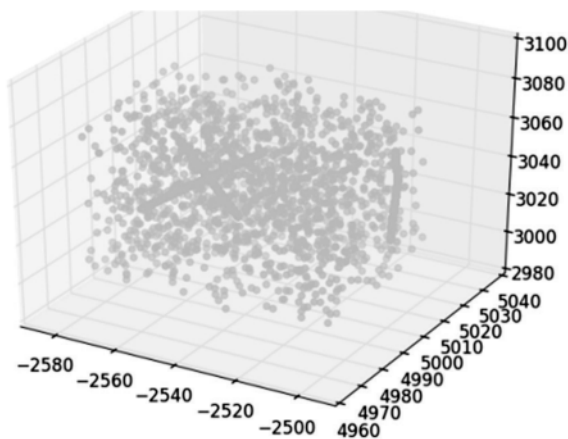


图 1 原始的航迹测试数据 3 维视图

2.4.2 聚类后数据

当选取 Eps 值为 5.522、11.251、17.191,MinPts 值为 2, ΔT 值为 0.04 s 时,分别进行聚类处理。得到的聚类结果如图 2~图 4 所示。

如图 2 所示,当取 Eps 值为 5.522,MinPts 值为 2, ΔT 值为 0.04 s 时,整个航迹点数据集被聚类为了 5 类(包括 4 类航迹和 1 类噪声),所有点迹的聚类正确率为 89.17%。

如图 3 所示,当取 Eps 值为 11.251,MinPts 值为 2, ΔT 值为 0.04 s 时,整个航迹点数据集被聚类为了 4 类(包括 3 类航迹和 1 类噪声),所有点迹的聚类正确率为 100%。

如图 4 所示,当取 Eps 值为 17.191,MinPts 值为 2, ΔT 值为 0.04 s 时,整个航迹点数据集被聚类为了 4 类(包括 3 类航迹和 1 类噪声),所有点迹的聚类正确率为 100%。

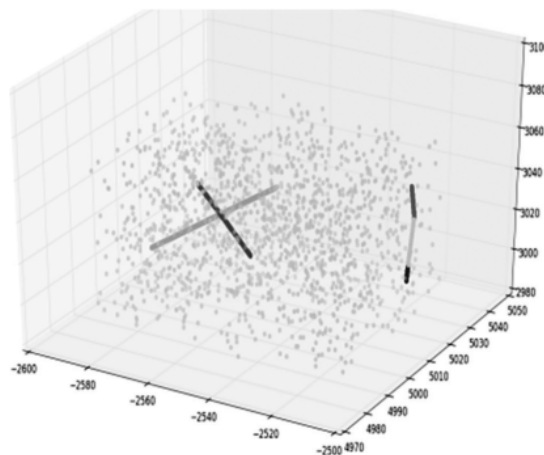


图 2 Eps 值为 5.522 时的聚类结果

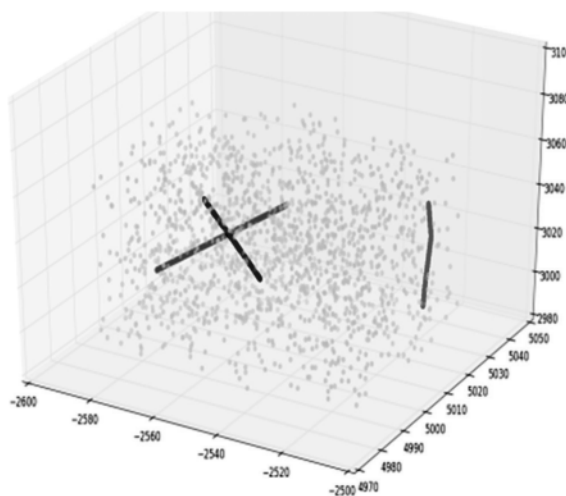


图 3 Eps 值为 11.251 时的聚类结果

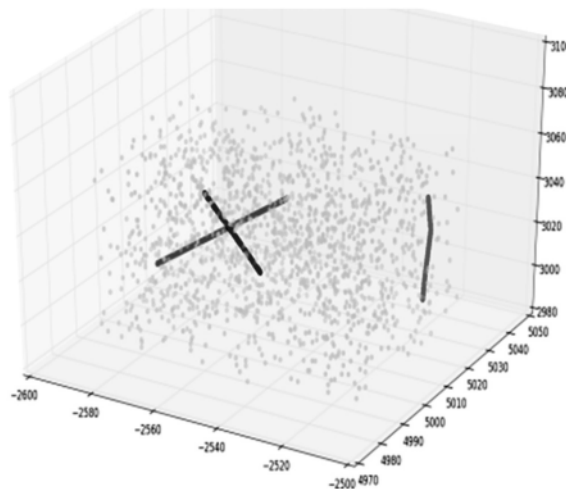


图 4 Eps 值为 17.191 时的聚类结果

2.4.3 聚类结果与分析

根据取 Eps 值为 5.522、11.251、17.191,MinPts 值为 2, ΔT 值为 0.04 s 时的 3 组参数的聚类效果分析可知,当以距离邻域 $Eps=5.522$ 为半径的圆球区域进行扫描时,

所有点迹的聚类正确率为 89.17%，其中第一类航迹点中有 4 484 个由于两点间欧式距离均超过了 5.523 因此被聚类为了噪声点，第二类航迹点中有 69 个由于两点间欧式距离均超过了 5.653 因此被聚类为了噪声点，其余点均聚类正确。

而当以距离邻域 $Eps=11.251$ 、 17.191 进行聚类时，以上航迹中的点均包含在了相应分类的核心点邻域从 Eps 为半径的圆球区域内，因此都被正确地聚类，聚类正确率为 100%。

3 结论

本文使用一种基于 DBSCAN 算法的 ST DBSCAN 算法进行了航迹聚类实现，ST DBSCAN 算法既符合 DBSCAN 的特点和优点又增加了从时间域对目标数据集的扫描。可对任意形状的航迹聚类，且不需提前划分聚类目标个数，并增加了对包含时空域航迹数据的聚类。通过增加对扫描邻域的时间连续性约束，从而从时间域维度提高了飞机航迹聚类的准确性。通过本文的仿真实验情况，在参数选择合适的条件下可以获得很好的聚类结果。但本方法也有和 DBSCAN 相同的不足之处，即对时空域上密度不均匀、密度差异大的点迹数据集的聚类效果不稳定，且聚类效果依赖对参数的选择等。在未来的工作中可以对此方法进行进一步的优化和完善。

参考文献

- [1] 宋金玉, 郭一平, 王斌. DBSCAN 聚类算法的参数配置方法研究[J]. 计算机技术与发展, 2019, 29(5): 44-48.
- [2] 熊回香, 叶佳鑫, 蒋武轩. 改进的 DBSCAN 聚类算法在社会化标注中的应用[J]. 数据分析与知识发现, 2018, 24(12): 77-87.
- [3] 李双庆, 慕升弟. 一种改进的 DBSCAN 算法及其应用[J]. 计算机工程与应用, 2014, 50(8): 72-76.
- [4] 刘世平. 数据挖掘技术及应用[M]. 北京: 高等教育出版社, 2010.
- [5] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48-61.
- [6] 邵峰晶, 于忠清, 王金龙. 数据挖掘原理与算法[M]. 北京: 科学出版社, 2009.
- [7] 张良, 陶海军, 杨钊, 等. 模糊聚类的多雷达航迹关联算法[J]. 现代防御技术, 2017, 45(6): 113-117.
- [8] 何爱林, 周德超, 陈萍, 等. 基于轨迹聚类的运动趋势分析[J]. 海军工程大学学报, 2017, 29(5): 103-107.
- [9] FRANKR. Clustering of flight tracks[C]. Proceedings of the American Institute of Aeronautics and Astronautics. Atlanta: AIAA, 2010.
- [10] 韩家炜, Micheline Kamber. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2007.
- [11] 韩家炜, 坎伯. 数据挖掘: 概念与技术英文版[M]. 3 版. 北京: 机械工业出版社, 2012.
- [12] BOWLES M. PYTHON 机器学习预测分析核心算法[M]. 沙赢, 李鹏, 译. 北京: 人民邮电出版社, 2017.
- [13] 张冉, 夏厚培. 一种新的 k-means 聚类雷达信号分选算法[J]. 现代防御技术, 2015, 43(6): 136-141.
- [14] 黄静, 官易楠. 基于改进 DBSCAN 算法的异常数据处理[J]. 软件导刊, 2020, 19(4): 219-223.
- [15] 刘勇. 基于 DBSCAN 的空间聚类算法研究与实现[D]. 昆明: 云南大学, 2017.

(收稿日期: 2021-11-15)

作者简介:

朱瑜亮(1984-), 男, 硕士, 工程师, 主要研究方向: 航空电子。



扫码下载电子文档

(上接第 124 页)

会, 2017: 643-646.

- [13] 刘巍巍, 谭承, 梁栋, 等. 基于 BGA 技术的毫米波垂直互连设计[C]//中国电子学会. 2017 年全国微波毫米波会议论文集(中册), 2019: 1522-1525.

(收稿日期: 2022-02-27)

作者简介:

刘昊东(1997-), 男, 硕士研究生, 主要研究方向: 小型化微波组件研制。

吴洪江(1964-), 男, 本科, 研究员, 主要研究方向: 固态微波器件和电路研制。

余小辉(1983-), 男, 硕士, 高工, 主要研究方向: 固态微波电路设计。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所