基于时空关联学习的轨迹频繁模式  
识别研究

吴施楷[[[1]](#footnote-0)],杜红林[1]

1. 成都傅立叶电子科技有限公司, 成都 610041)

**摘 要:** 随着雷达定位技术的日益完善,多源轨迹数据的获取也越来越容易，这些繁杂的轨迹数据中蕴含着大量的航行运动属性及航行意图信息, 对这些轨迹数据进行深度挖掘变得至关重要。目标对象多源轨迹数据的频繁模式是在轨迹数据融合的基础上开展的二次信息挖掘，目的是生成目标航行的轨迹模板，提供目标类型识别及异常航行行为识别依据。本文基于时空关联序列模式挖掘的思路，在常规频繁项识别算法的基础上结合密度峰值聚类算法实现了时空关联的频繁模式提取算法，该算法首先对目标的多源轨迹集进行聚类编码，其次将编码序列按时间域进行拼接，最后采用改进PrefixSpan算法识别其中连续出现的频繁模式。文中利用真实轨迹数据进行了实验分析, 结果表明, 所提方法能有效融合复杂轨迹类型。与常规频繁项识别算法相比，融合相似度明显提高，轨迹相似度可达到93%以上。

**关键词: 多源轨迹;频繁模式;数据融合;时空关联;密度峰值聚类;编码序列;PrefixSpan**

**Research on Trajectory Frequent Pattern Recognition**

**based on Spatio-Temporal Correlation Learning**

**Wu shi-kai[1], Du hong-lin[1]**

1. **Chengdu Fourier Electronic Technology Co., LTD 610041)**

**Abstract:** With the improvement of geographic location information acquisition technologies such as radar, it has become easier to acquire multi-source spatio-temporal trajectory data. These complex trajectory data contain a large amount of navigational motion attributes and navigation intent information. Deep digging becomes essential. The frequent pattern of the target object's multi-source trajectory data is secondary mining on the basis of trajectory data fusion. The purpose is to generate the trajectory template of the target navigation, and provide the basis for target type recognition and abnormal navigation behavior recognition. Based on the idea of mining spatio-temporal correlation sequence patterns, this paper combines the density peak clustering algorithm on the basis of the conventional frequent item recognition algorithm to realize the spatiotemporal correlation frequent pattern extraction algorithm. The algorithm firstly clusters and encodes the target's multi-source trajectory. Secondly, the coding sequence is spliced in the time domain, and finally the PrefixSpan algorithm is used to identify the frequent patterns that appear continuously. In this paper, the real track data are used for experimental analysis. The results show that the proposed method can effectively integrate complex trajectory types. Compared with the conventional frequent item recognition algorithm, the fusion acquaintance is significantly improved, and the trajectory similarity is significantly improved. Reached 93%.

**Keywords:** Multi-source Trajectory; Frequent Pattern; Data Fusion; Temporal and Spatial Correlation; Density Peak Clustering; Coding Sequence;PrefixSpan

# 1 引言

随着雷达技术和全球定位系统的不断发展, 越来越多的轨迹数据能够被收集到[1]，目标轨迹数据是目标最直接、最基础的信息数据，既包含了目标实时位置信息，也蕴含了目标的基本运动属性，是构建目标识别、位置预测、意图分析等模型的基础属性特征[2][3]。针对多源轨迹数据，首要任务是实现实时轨迹数据关联融合，利用历史数据融合构建的频繁轨迹模板进行目标类型识别及异常行为识别[4][5]。特别是在现代雷达侦察场景下，对于敌方的侦察、战势的推断、威胁的评估等都需要快速、精确地掌握目标的基础信息[6]，充分利用大量历史航迹数据挖掘先验知识是进行目标实时分析的关键。近年来，受雷达侦察应用的驱动，涌现出大量轨迹挖掘模式（包括序列模式、周期模式、伴随模式、频繁模式等）。序列模式可以用于预测目标对象下一个航经的位置[7],在周期模式挖掘中，Li等[8]提出分两步发现关于某些地点的周期模式，即首先使用傅立叶变换和自相关函数找到其中的周期，之后对不同时间的周期行为进行层次聚类。Zheng等[9]使用聚类算法和Hausdorff距离，挖掘一类特殊的伴随模式，实现多目标的编队行为识别。Song等[10]对民航航班轨迹数据进行了重采样，然后使用主成分分析和DBSCAN算法进行长轨迹聚类，最后将得到的聚类中心作为轨迹频繁模板，用于改进已有的轨迹识别。

传统的频繁项挖掘识别主要是针对离散项，算法多采用Apriori算法[11]，FP-growth算法[12]，PrefixSpan算法[13][14]。这类算法通过统计的方式寻找样本集空间中的关联属性，但是，现实情况下，连续时空也存在很多的关联性，这些关联性可以作为后续的任务规划及任务调度的先验知识。尽管后来在Aprior算法上衍生出来的PrefixSpan可以解决连续空间频繁项发现，但是，这种频繁项只找出了空间上的关联性，缺失了时间关联性，表现出时间上的离散跳跃，而且对单一轨迹无法提取频繁子段。因此，针对扩展雷达侦察理论知识及对频繁挖掘算法的提炼及改进，本文提出了一种基于连续时间、空间的频繁项挖掘及融合算法，并在指定轨迹集上进行了算法验证。

# 2 频繁模式识别理论

## 2.1 轨迹聚类预处理

运动轨迹的分布模式是目标行为特征的重要表现方式之一，在许多目标侦察场合中通过对目标运动轨迹的分析就可以提取出目标的行为模式。然而在多源轨迹数据融合之前，需要将数据进行分组、异常数据剔除及轨迹连续光滑[15]。在没有标签数据引导情况下，需要采用无监督聚类算法进行数据预处理。目前国内外对此研究较深入的有基于轨迹之间不同距离度量方式的聚类分析方法[16]、基于轨迹特征表示的聚类分析方法[17]和基于轨迹分布建模的聚类分析方法[18][19]等。这些方法大都需要度量不同轨迹之间的相似性并据此进行聚类。然而，不同目标的运动轨迹在长度和运动速度等方面都有差别，很难以一种非常自然的方式来度量它们之间的相似性。另外，当场景中目标的运动比较复杂时，很难对它们进行准确的聚类，聚类数目的确定也比较困难。这种随机性的运动也会影响到基于聚类的轨迹分析方法的效果。基于距离及密度的聚类算法很难在复杂轨迹模式下确定距离阈值及分类个数，因此本文采用基于密度顶点及距离联合的密度峰值聚类算法[20]。

该算法假设聚类中心由那些带小的局部密度值的临近点包围，且它们与那些具有较大局部密度值的点有相对大的距离。对于每一个点，计算两个值：它的局部密度值，与较高密度值点的距离。这两个属性值仅仅决定于点与点之间距离，距离满足三角不等式。局部密度定义为

 （1-1）

其中，如果,,否则。是截断距离。可以看出等于那些距离接近点的截断距离的点的数量。该算法唯一地对在不同点的的量级敏感，也就是说，对于大的数据集，分析的结果关于的选择具有鲁棒性。

是点与那些具有较高密度点之间距离的最小值：

 （1-2）

具有最大密度的点，。注意只有那些密度是局部或者全局最大的点才会有远大于正常的相邻点间距。

那些有着比较大的局部密度和很大的的点被认为是簇的中心。局部密度较小但是较大的点是异常点。在确定了簇中心之后, 所有其他点属于距离其最近的簇中心所代表的类簇。

## 2.2 连续频繁项识别

假设数据集的存在一个项集的支撑，表示为，即中所有包含的事务的数目为

（1-3）

的相对支撑为

（1-4）

它是对包含的项的联合概率密度的估计。若，则称项集在中是频繁的，其中是给定的最小支撑阈值。

一条规则的置信度是一个事务包含的情况下也包含的条件概率：

（1-5）

当是一条强规则。

由于轨迹反映的是目标在监控场景中连续运动时的规律，因此在轨迹序列模式的挖掘中，鉴于提出的PrefixSpan（Prefix projected Sequential pattern mining）算法是目前序列模式挖掘中性能较好的一种方法，本文对其进行了改进，使其能够用于连续的轨迹序列模式挖掘。基于改进PrefixSpan 的轨迹模式挖掘算法首先通过流矢量量化编码将轨迹变换成符号序列的集合，然后采PrefixSpan 投影挖掘的思想找出其中频繁出现的序列模式作为轨迹分布的表示[21]。

PrefixSpan提取出的频繁序列在时间序列上存在跳跃，提取出的频繁项不具备时间上的连续性，因此，本文在PrefixSpan的基础上加入时间连续性，提取出的频繁项能更好地匹配真实数据。



图 1 频繁轨迹生成流程

# 3 仿真实验与分析

## 3.1 仿真序列数据验证

为了对比常规PrefixSpan算法与改进PrefixSpan算法在连续序列上的效果，本文采用了单序列测试样例与多序列测试样例，多测试样例以二维代表，横轴代表时间，纵轴代表一维空间。单序列样例主要为了测试时间序列上的多次出现项。

多维时序数据测试样例为S = [[4,3,4,0],[0,1,3,4,3,4,3,4,0,1]]。PrefixSpan计算结果如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4 | 43 | 434 | 4340 | 430 | 44 | 440 | 40 | 3 | 34 | 340 | 30 | 0 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

表 1 PrefixSpan计算结果（第一行为计算序列，第二行为频繁次数）

改进PrefixSpan计算结果如下：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 0 | 01 | 34 | 340 | 4 | 40 | 43 | 4340 |
| 2 | 3 | 2 | 4 | 2 | 5 | 2 | 3 | 2 |

表 2 改进PrefixSpan计算结果（第一行为计算序列，第二行为频繁次数）

单序列测试样例为S = [0,1,3,4,3,4,3,4,0,1]

PrefixSpan计算结果为空，没有计算结果，说明该算法对单时间序列不具备应用价值。改进PrefixSpan计算结果如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 01 | 4 | 34 | 43 |
| 2 | 2 | 3 | 3 | 2 |

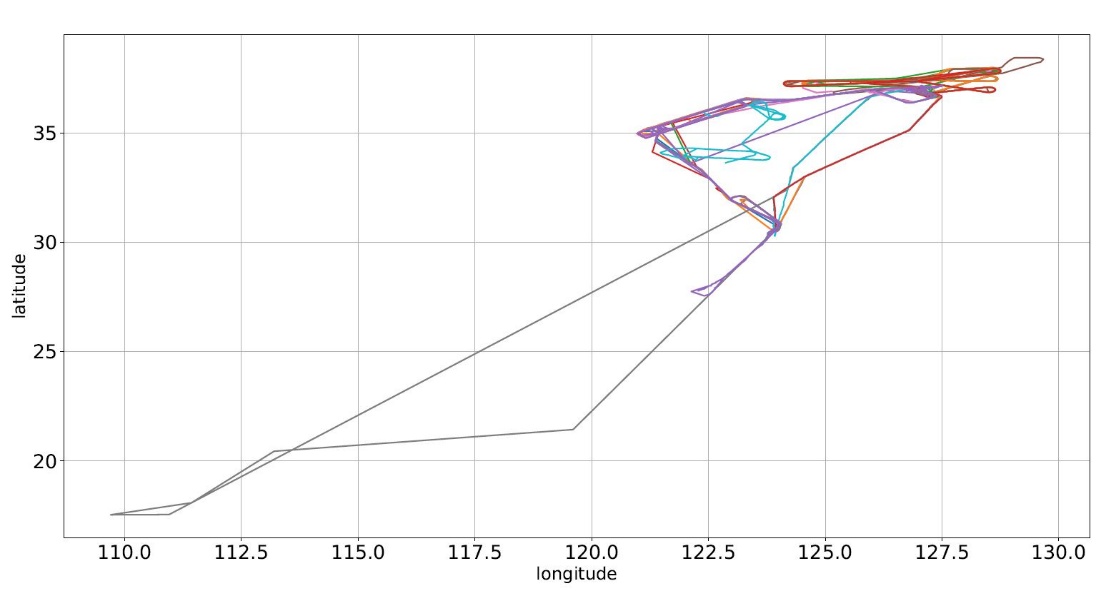
表 3 改进PrefixSpan计算结果

表一、表二、表三对比几点不同主要是：

1. PrefixSpan计算频繁项明显较多；
2. 计算出的频繁项在原序列中不连续、较离散，比如440、44、30等在原序列中是间断的；
3. 频繁次数明显较少，比如34、43等
4. 对单序列常规PrefixSpan无法计算结果

## 3.2 实测轨迹数据验证

实测数据采用ADS-B[22][23]原始数据集，分别来自同一目标不同时间、不同区域的轨迹数据，原始数据可视化见下图，可以看出同一目标飞行轨迹形状多样，分布区域广，盘旋次数多，并且存在多条异常数据。可视化分析得出轨迹族群比较明显，主要集中在两个区域1、2，每个区域中同时存在多个子区域。但是，各子区域航线不规整，航道交叉偏离，将对后续航线预测及意图分析产生较大误差。因此，需要在大族的基础上进一步挖掘子族，以及对子族进行合并融合。



4111

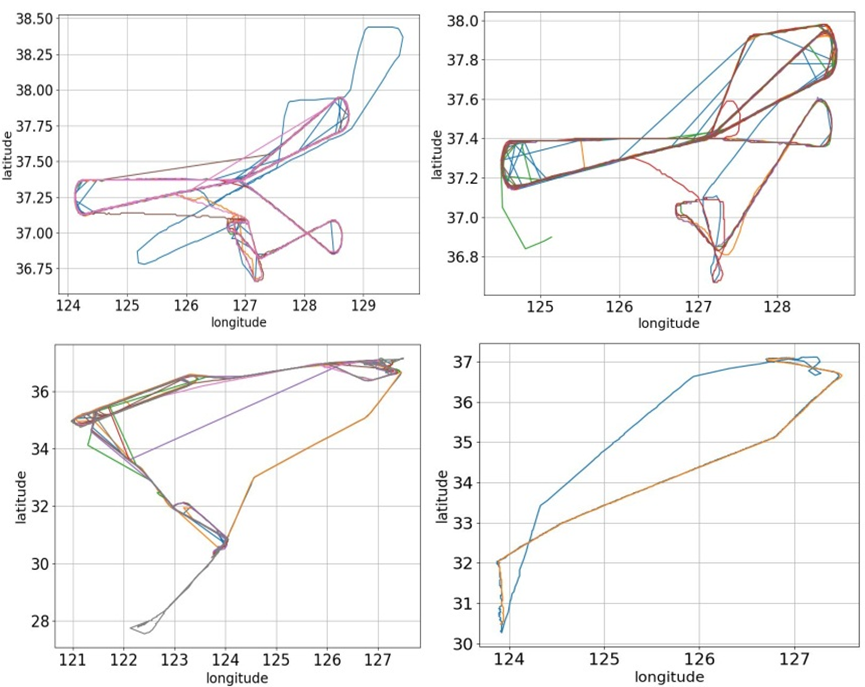
3111

2111

12111

图 2 原始轨迹数据二维可视化

采用密度峰值聚类算法对原始数据进行主要类族划分，结果如图3、图4，其中图三3表示原始数据的主要分布，图4表示异常轨迹。可以看出，轨迹族共分类四类，其中1代表图2中的类族1，2代表图2中的类族2。通过聚类分组之后，轨迹线路划分更精细，可以明显捕捉到航线的方向、大曲率转角及多次盘旋，这些特征对航行目标分类分析以及飞行意图分析提供直接依据。但是，航线族内航线段差异大，噪声干扰严重，主要原因是定位误差以及采样点丢失，图中以箭头指向表示。因此，为保证主航线的精确定位及异常航线段的剔除，需要计算出航线的频繁轨迹段，并按照时间连续性进行频繁轨迹段的拼接融合。



2111

2111

12111

12111

图 3 轨迹数据聚类粗划分主要类族

异常航迹主要表现为两类，1.跨度范围超出主航线实际区域；2.航线次数少。第一类是错误数据，不作为分析对象。产生第二类的主要原因有两类，数据采样少及航向目标的异常航行行为，这类数据可作为异常航线的识别依据。

4111

3111

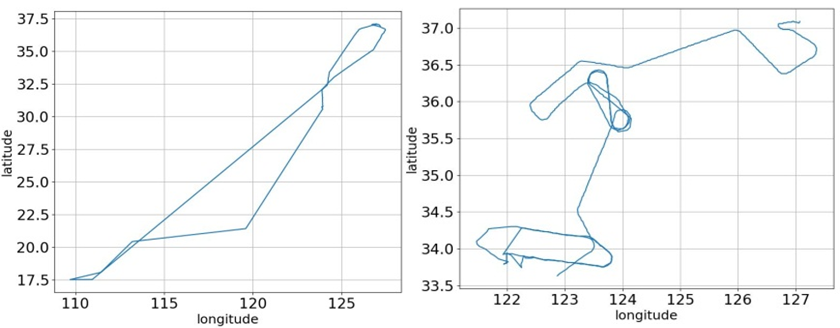
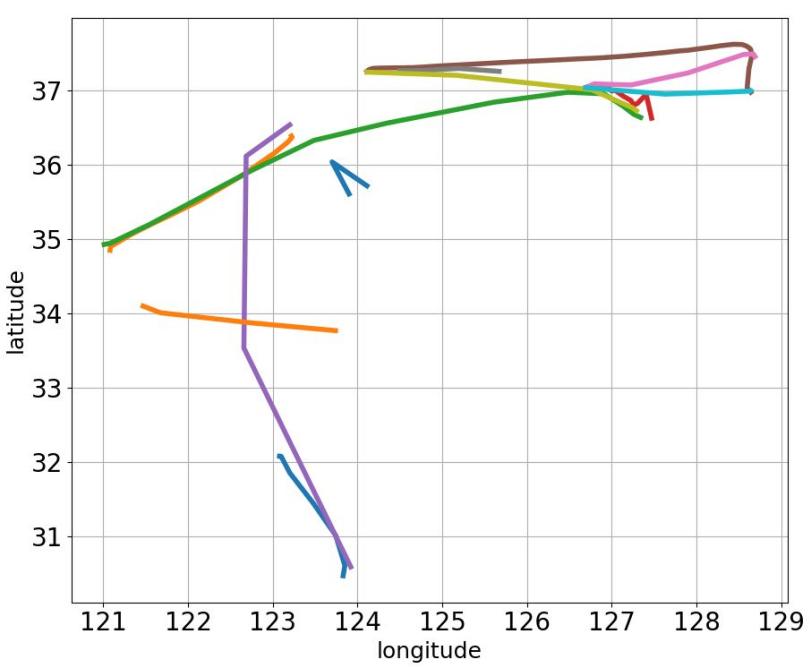


图 4 识别异常轨迹

为了实现主航线的精确定位及异常航线段的剔除，需要计算轨迹的频繁轨迹段及时序融合，本文采用PrefixSpan算法与改进频繁算法实现数据的融合，实际轨迹数据测试结果可视化见图5、图6、图7，为量化对比效果，本文将频繁轨迹段与原始数据进行比对分析，计算出轨迹之间的比对相似度。



离散段

转弯段

直线段

图 5 常规PrefixSpan处理的频繁轨迹段

从上图可以看出，常规PrefixSpan算法处理后的融合轨迹分段较多，缺失空间连续性。该算法对轨迹直线段捕捉较为准确，能实现一定范围转弯半径的轨迹段识别，对连续转弯段及盘旋段无法准确识别，丢失了轨迹原有的形态信息，无法形成航线先验知识。实际的侦察任务中，迫切需要对连续转弯、多次盘旋的位置进行精确的定位，保证时空上的连续性及航线的可追踪性、可预测性。

本文算法计算结果见图6、图7，转弯、盘旋段可以被精确识别，异常航迹段被过滤，数据由多条融合为一条完整轨迹，轨迹段光滑连续拼接，不存在断点、离散问题。但是，某些局部轨迹段被过度光滑，细节信息存在丢失，分析主要原因是轨迹整体异常段多、轨迹盘旋区域经纬度方差大，频繁算法为保证轨迹全局平滑性而造成局部段过度拟合。

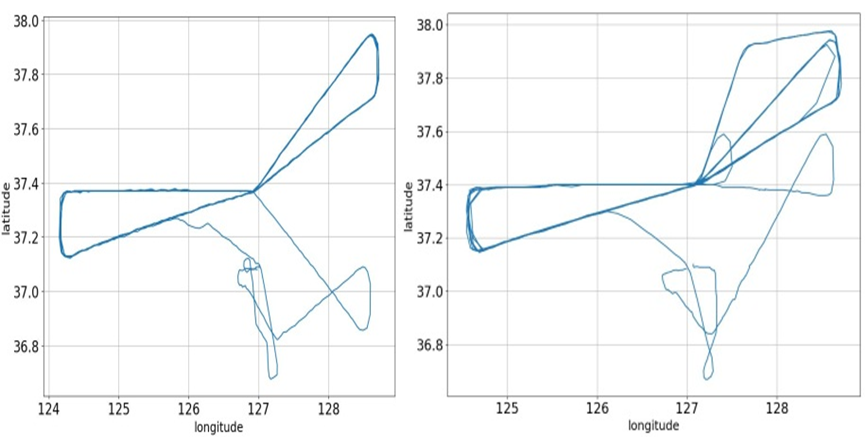
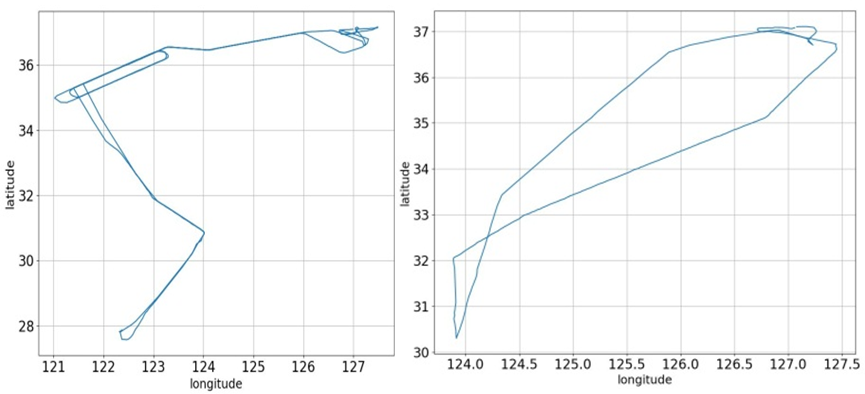


图 6 改进PrefixSpan算法第一组频繁航线



过拟合段

图 7 改进PrefixSpan算法第二组频繁航线

以上通过可视化方式阐释了时空关联学习的频繁轨迹识别算法的实际验证结果，为对比算法差异性，需要定义量化指标，本文采用融合频繁轨迹与原始轨迹之间的相似度作为算法效果指标，相似度计算公式采用轨迹Hausdorff距离[24][25]公式。常规算法与改进算法频繁轨迹比对相似度如下，从表4数据计算出改进算法频繁轨迹平均相似度达到93.25%，满足实际使用的预期效果。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | 第一组相似度  （%） | 第二组相似度  （%） | 第三组相似度  （%） | 第四组相似度  （%） |
| 常规算法 | 77 | 70 | 82 | 79 |
| 改进算法 | 92 | 89 | 97 | 95 |

表 4 轨迹比对量化相似度

# 4 结束语

常规多源轨迹数据频繁模式识别算法缺乏时空关联，生成的频繁轨迹对复杂盘旋轨迹段不敏感，无法识别。本文基于常规频繁模式识别算法，结合密度峰值聚类算法实现了时空关联的频繁模式提取算法，该算法充分考虑数据多源性以及时空连续性，提高了频繁轨迹挖掘效率及轨迹相似度，同时本算法具备一定的抗噪声干扰能力。但本算法仍然存在多干扰条件下的过拟合问题，接下来的研究中会重点关注并优化算法。本研究可应用于雷达侦察领域数据规模大、轨迹模式难以确定场合下的敏感目标识别以及实时比对侦察目标异常行为等。

# 参考文献

[1] 宋嘉庚，张扶桑等.基于频繁航路模式的航迹类型识别[J].计算机科学，2021,48(9):59-67.

[2] 徐建平，吴蔚等.多雷达信息融合系统中的异常轨迹检测[C].第四届中国指挥控制大会论文集，2016:602-606.

[3] 史建涛, 孙俊, 杨予昊,等. 基于机器学习支持向量机的海面目标航迹起始算法研究及应用[J]. 现代雷达, 2019(11):20-24.

[4] 鹿强，吴琳等.海上目标的多源轨迹数据关联综述[J].地球信息科学学报，2018,20(005):571-581.

[5] 韩昭蓉, 黄廷磊, 任文娟,等. 基于Bi-LSTM模型的轨迹异常点检测算法[J]. 雷达学报, 2019, 8(01):40-47.

[6] Ping Lang, Xiongjun Fu. A Comprehensive Survey of Machine Learning Applied to Radar Signal Processing[J].2020.

[7] MONREALE A, PINELLI F, et al. WhereNext: A location predictor on trajectory pattern mining[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France: ACM, 2009.

[8] LI Z,DING B, et al. Mining periodic behaviors for moving objects[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.2010:1099-1108.

[9] ZHENG K, ZHENG Y, et al. On Discovery of gathering patterns from trajectories[C]//2013 IEEE 29th International Conference on Data Engineering. IEEE, 2013:242-253.

[10] SONG Y, CHENG P, MU C. An improved trajectory prediction algorithm based on trajectory data mining for air traffic management[C]//2012 IEEE international Conference on Information and Automation. IEEE, 2012:981-986.

[11] 王成勇.关联规则Aprior算法的研究与应用[D].华北电力大学，2019.

[12] 晏杰.基于Aprior&FP-growth算法的研究[J].计算机系统应用，2013,005(2013):122-125.

[13] Wang LL, Fan J. Improved Algorithm for sequential pattern Mining Based on PrefixSpan [J].Computer Engineering, 2009.

[14] 刘松,朱文佳，罗达志. 一种基于PrefixSpan算法的车辆频繁轨迹挖掘方法及系统[P]，CN111009123A.2020.

[15] 李宏博, 沈一鹰. 一种基于航迹光滑滤波的目标跟踪方法[J]. 现代雷达, 2009, 31(3):43-47.

[16] PAN J, JIANG Q, SHAO Z. Trajectory clustering by sampling and density[J].Marine Technology Society Journal,2014, 48(6):74-85.

[17] Guan Yuan, Shixiong Xia. An efficient trajectory-clustering algorihm based on an index tree[J]. Transations of the Institute of Measurement and Control, 2011:1-12.

[18] 高强，张凤荔，王瑞锦，等．轨迹大数据: 数据处理关键技术研究综述［J］．软件学报，2017，28( 4) :8-10．

[19] 冯振华，钱雪忠，赵娜娜．GreedyDBSCAN:一种针对多密度聚类的DBSCAN改进算法［J］.计算机应用研究,2016，33( 9) :2-4．

[20] Rodriguez A, Liao A. Clustering by fast search and find of density peaks[J].Science, 2014,344(6191):1492.

[21] Pei J. PrefixSpan: Mining Sequential patterns efficiently by Prefix-Projected pattern growth[J]. Proc.2001 Int. Conf. Data Engineering, 2001.

[22] 张瞩熹, 田旺, 朱少川,等. 多源异构航班航迹数据流实时融合方法研究[J]. 物联网学报, 2020, 4(3):60-68.

[23] 马兰, 高永胜. 基于ADS-B数据挖掘的4D航迹预测方法[J]. 中国民航大学学报, 2019, 037(004):1-4.

[24] 张晓滨, 杨东山. 基于时间约束的Hausdorff距离的时空轨迹相似度量[J]. 计算机应用研究, 2017, 34(007):2077-2079.

[25] 黄天镜. 基于多维特征的ADS-B数据异常检测方法研究[D]. 中国民航大学, 2020.

# 作者简介

吴施楷，男，1989，硕士，通信侦察与人工智能.(E-mail:tianqin8990@163.com)

**\***通信作者 杜红林，男，1987，本科，通信侦察. (18140038015，duhl@sdgi.com.cn，四川省成都市武侯区武兴二路17号力德时代12栋A座6层 成都傅立叶电子科技有限公司)

1. 作者简介 吴施楷，男，1989，硕士，通信侦察与人工智能. ( E-mail: tianqin8990@ 163.com)

   通讯作者 杜红林，男，1987，本科，通信侦察. ( E-mail: duhl@sdgi.com.cn) [↑](#footnote-ref-0)