

基于渔船轨迹数据的进出港区域识别方法

崔彤彤,徐 硕,刘慧媛

(中国水产科学研究院 渔业工程研究所,北京 100141)

摘要:针对当前渔船进出港区域获取方法成本高、更新周期长等问题,提出了一种基于渔船轨迹数据的渔船进出港区域识别方法。首先,提出基于多特征融合下轨迹点间相似性的轨迹划分算法,将渔船轨迹划分为不同渔船行为的轨迹段;然后,提出特征距离加权-K均值聚类算法(Feature Distance Weighted-K-means clustering algorithm, FDW-K-means),将上一步得到的轨迹段特征作为聚类对象,实现渔船进出港轨迹段的提取。最后,综合运用DBSCAN(Density-Based Spatial Clustering of Applications with Noise)聚类算法和Del-Alpha-Shape算法对聚集的渔船进出港轨迹段轨迹点集进行边界提取获得渔船进出港区域。以椒江渔港和博贺渔港2021年3月的渔船轨迹数据为例,识别到椒江渔港和博贺渔港的渔船进出港区域的正确率分别为94.2%和95.8%。与使用K-means聚类算法或传统基于对各特征设定约束条件思想提取轨迹段的方法相比,该方法识别到的渔港渔船进出港区域正确率分别提高了10.7%、8.7%和9.5%、6.6%。实验结果表明所提方法能够有效识别渔船进出港区域,其结果能为渔船进出港监管提供科学参考。

关键词:渔船轨迹数据;多特征融合;轨迹划分;K-means;进出港轨迹段;进出港区域

中图分类号:TP399

文献标识码:A

文章编号:1673-629X(2024)06-0155-09

doi:10.20165/j.cnki.ISSN1673-629X.2024.0068

Identification Method of Port Entering and Leaving Area Based on Fishing Vessel Trajectory Data

CUI Tong-tong, XU Shuo, LIU Hui-yuan

(Institute of Fisheries Engineering, Chinese Academy of Fishery Sciences, Beijing 100141, China)

Abstract: In view of the problems such as high acquisition cost and long update period, a new method based on fishing vessel trajectory data was proposed to identify port fishing vessel entering and exiting area. Firstly, a trajectory partitioning algorithm based on the similarity between trajectory points under multi-feature fusion was proposed. The fishing vessel trajectory data was divided into trajectory subsegments with different behaviors. Then, the Feature Distance Weighted-K-means clustering algorithm (FDW-K-means) was proposed to cluster the features of trajectory segment, and the trajectory segments of fishing vessel entering and leaving were obtained. Finally, the DBSCAN clustering algorithm and Del-Alpha-Shape algorithm were used to extract the boundary of the gathered data set of the entering and leaving trajectory segment of fishing vessel to obtain the port fishing vessel entering and leaving area. Based on the fishing vessel trajectory data collected in March 2021 at Jiaojiang Fishing port and Bohe fishing port, the accuracy of fishing vessel entering and exiting area was 94.2% and 95.8%, respectively. Compared with K-means or the traditional method based on the idea of setting the threshold value of features, the accuracy of extracting the entering and exiting area of fishing vessel in the two fishing ports was increased by 10.7%, 8.7% and 9.5%, 6.6%, respectively. The experimental result shows that the proposed method can effectively extract the port entering and leaving area of fishing vessel, which can provide scientific reference for the supervision of fishing vessel entering and leaving the port.

Key words: fishing vessel trajectory data; multi-feature fusion; trajectory partitioning; K-means; entering and leaving trajectory segment; port entering and leaving area

0 引言

渔船进出港区域是位于渔港海岸线附近,众多渔

船进入渔港停靠前或驶离渔港时通常会经过的水上区域,一般沿海岸线分布^[1]。渔船进出港区域对掌握渔

收稿日期:2023-08-09

修回日期:2023-12-13

基金项目:国家农业科学数据中心渔业科学数据资源建设与共享服务(NASDC2023XM03);中国水产科学研究院渔业工程研究所基本科研业务费专项(2022HY-ZC004);渔业通信导航与大数据创新团队项目(2020TD84)

作者简介:崔彤彤(1993-),女(蒙古族),硕士,助理研究员,研究方向为渔业信息化;通讯作者:徐 硕(1985-),男,硕士,副研究员,研究方向为渔业信息化。

船进出港情况,加强渔船进出港监管具有重要意义。当前采用人工或利用遥感影像等方式来采集渔船进出港区域,但该方法消耗大量人力物力,成本高,更新周期长,难以满足实际的应用需求。

近年来,各类轨迹数据采集系统的兴起使得渔船在海面上航行产生了大量的渔船轨迹数据,渔船轨迹数据逐渐被推广应用到渔船行为分析、渔船管理的研究领域。具体反映到识别渔船进出港区域研究方面,渔船进出港时与航行在其他地理位置有不同的行为状态,使得经过渔船进出港区域的渔船轨迹数据表现出不同于经过其他区域的航行特征。通过众多渔船进出港时的轨迹数据来识别到渔船进出港区域逐渐成为可能。船舶自动识别系统(Automatic Identification System, AIS)是目前使用较为广泛的开放式数据传输系统^[2],数据采集便利、更新快,有力地推动了航运智能化发展,并在船舶航行时空规律、轨迹预测、港口地理学等研究中发挥了重要作用^[3]。通过信息化技术基于 AIS 采集到的渔船轨迹数据识别渔船进出港区域,如果可行还具有成本低、更新快等特点。但是,采集的渔船轨迹数据具有渔船行为难以直接表现等特点^[4],如何从大量的渔船轨迹数据中提取到符合渔船进出港区域的轨迹数据是识别渔船进出港区域的关键。因此,该文提出了基于渔船轨迹数据的渔船进出港区域识别方法。具体贡献包括:

(1)针对渔船轨迹数据难以直接表现不同渔船行为的问题,提出了一种基于多特征融合下轨迹点间相似性的轨迹划分算法,将渔船轨迹划分为不同渔船行为的轨迹段,为下一步提取具有渔船进出港行为的轨迹段提供有效的数据支持;

(2)提出了特征距离加权-K-means 聚类算法(Feature Distance Weighted-K-means clustering algorithm, FDW-K-means),将划分后的轨迹段的特征作为聚类对象,实现相同行为轨迹段的聚合,得到具有渔船进出港行为的轨迹段;

(3)综合运用 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)聚类算法和 Delauney 三角剖分算法与 Alpha Shapes 轮廓线提取算法相结合的 Del-Alpha-Shape 算法^[5],对聚集的渔船进出港轨迹段进行边界提取,得到渔船进出港区域。

1 相关工作

当前,基于渔船轨迹数据进行渔船进出港区域识别的直接相关研究还比较少,国内外学者围绕陆地交通轨迹信息或其他船舶轨迹信息研究有一些相关工作。

在具有某种轨迹行为的轨迹段提取方面,主要分

为利用地理背景信息的方法、挖掘轨迹特征的方法和利用聚类算法对轨迹数据直接聚类的方法。Alvares 等^[6]通过利用地理信息丰富移动对象轨迹数据,从而实现感兴趣轨迹点或轨迹段的分析和挖掘。Zheng 等^[7]从地理背景信息出发,提出了通过挖掘移动对象和已访问地点之间的相关性,计算轨迹在该地理位置或其邻域范围内的连续累积时间来判断轨迹在该处的行为,从而提取轨迹段。但是上述两种方法过于依赖地理背景信息,在信息存在缺失时会对结果产生一定的负面影响。Rocha 等^[8]使用基于轨迹数据的方向等特定特征,并在特征上设置一定的规则来获取需要的轨迹信息。但是该类方法的特征选择和规则设置需要提前考虑到是否对轨迹段提取有决定作用,在通用性上会有一定局限。郑振涛等^[9]基于轨迹的特征提出了一种多约束条件下停留轨迹段提取方法,但是该方法依赖于每一个特征的约束条件阈值,阈值过大或过小都会造成提取到的轨迹数据包含更多的噪声点,影响提取结果。Palma 等^[10]提出了一种基于轨迹点速度的时空聚类方法来获取特定轨迹数据集,但是该方法要求具有大量高精度位置数据,而不适合渔船 AIS 轨迹数据质量低、采样稀疏的情况。张治华等^[11]提出了一种逐级合并的方法来提取特定的轨迹数据,该方法虽然在一定程度上克服了噪声数据的影响,但是其简单的合并规则易导致一些相似但不同的行为轨迹数据的合并。K-means 算法是一种经典的基于划分的无监督学习算法,近年来在轨迹行为聚类方面被广泛应用^[12]。耿家利^[13]、吴宝福^[4]使用 K-means 算法提出了基于轨迹特征的多步聚类方法实现渔船轨迹行为的判断和提取,但是二者在轨迹段聚类时并没有考虑特征的重要程度,对聚类的效果具有一定影响。由于使用传统的 K-means 算法对轨迹特征表达不全面,王代楠等^[14]提出了一种引进差分进化算法来改进 K-means 算法挖掘船舶航行在密集位置的轨迹段,但结论认为该方法限制于参数设定,不同参数设置对算法执行的有效性、复杂度和资源占用率的影响也有一定区别。

在区域识别方面,当前学者围绕功能区域识别展开的研究工作主要通过一定约束条件构造的数据集提取边界来实现。杨伟等^[15]通过采用 Delaunay 三角剖分算法对众多车辆停留轨迹数据集提取边界实现加油站区域的识别;惠振阳等^[16]采用 Alpha Shapes 轮廓线提取算法获得建筑物轮廓线。但从内容上看,上述方法无法直接适用于渔船轨迹数据的边界提取。

综上所述,基于众多渔船轨迹数据识别渔船进出港区域需解决两个问题:①如何从渔船轨迹数据中提取到具有渔船进出港行为的轨迹段;②如何根据具有

渔船进出港行为的轨迹段识别到渔船进出港区域。因此,该文针对已有研究方法的局限性,提出了一种渔船进出港区域识别方法。首先通过提出的多特征融合下轨迹点间相似性的轨迹划分算法与提出的 FDW-K-means 聚类算法,实现具有渔船进出港行为的轨迹段提取,然后综合运用 DBSCAN 聚类算法和 Del-Alpha-Shape 算法对聚集的渔船进出港轨迹段进行边界提取,最终得到渔船进出港区域。

2 基于渔船轨迹数据的渔船进出港区域识别有关概念

2.1 基础定义

定义 1 渔船轨迹。 渔船轨迹是一艘渔船一段时间范围内所在位置的时空序列。可被描述为: $T = (P_0, P_1, \dots, P_n)$, n 为该段时间内采集到的渔船轨迹点的个数, P 表示渔船轨迹的轨迹点,每个轨迹点包含船名、经纬度、时间戳、航速、航首向等属性信息,可描述为: $P = (\text{shipname}, \text{utc}, \text{lon}, \text{lat}, v, h)$ 。其中,航速单位是节,航首向是指以地理北为基准,顺时针到渔船在某一时刻船头正对方向之间的夹角,范围为 $[0^\circ, 360^\circ]$ 。

定义 2 渔船轨迹点集。 一系列轨迹点 P 构成的集合称为轨迹点集。表示为 $\text{TG} = \{P_0, P_1, \dots, P_m\}$,其中, m 为轨迹点集中轨迹点 P 的个数。

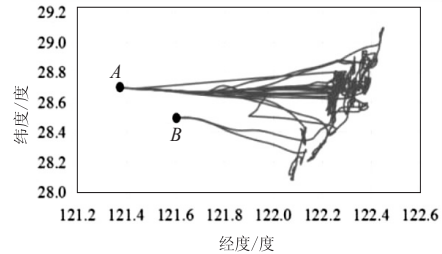
定义 3 渔船轨迹段。 渔船轨迹段是一艘渔船的一条渔船轨迹 T 中一段按照采集时间排序的轨迹。可描述为: $T_s = (P_i, P_{i+1}, \dots, P_{j-1}, P_j)$;其中, $P_i, P_{i+1}, \dots, P_{j-1}, P_j$ 为渔船轨迹段的轨迹点,按照采集时间顺序排序,且 $0 \leq i < j \leq n$, s 为渔船轨迹段序号。

2.2 渔船进出港区域相关概念

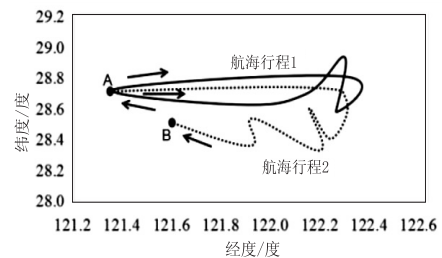
渔船行为通常包括停泊行为、作业行为和航行行为等。渔船进出港行为类似航行行为,不同于往返渔场之间和作业前后的航行行为,渔船进出港行为是渔船在一次航海行程中驶离渔港后和回港时停靠渔港前发生的行为。渔船进出港行为表现出一段进港轨迹或出港轨迹,由一些按照时间顺序连续的进港轨迹点或出港轨迹点组成,被称为渔船进港轨迹段或渔船出港轨迹段。

图 1(a)所示为一艘渔船在一段时间内多次出海作业的航海行程轨迹,点 A 与点 B 为渔船在渔港的泊位点,渔船一次航海行程与泊位点的关系有 2 种类型:①渔船出港前的泊位点与渔船出海作业返港后停泊的泊位点相同,如图 1(b)中的航海行程 1;②渔船出海前的泊位点与渔船出海作业返港后停泊的泊位点不同,如图 1(b)中的航海行程 2。不同渔船进港和出港可能会经过同一片区域,众多渔船的进港轨迹段和出港轨迹段在该区域会有聚集模式,具有相似的轨迹特

征,因此该文对渔船的进港轨迹段和出港轨迹段统一研究,将其统称为渔船进出港轨迹段。众多渔船进出港轨迹段经过的区域称为渔船进出港区域,可通过众多聚集渔船进出港轨迹段的轨迹点集边界来表示,是一个不规则区域^[9]。



(a) 渔船多次出海轨迹



(b) 渔船两次航海行程轨迹

图 1 某渔船出海作业轨迹

渔船轨迹点的特征属性包括航速、航首向、加速度、曲率等,其中,航速和航首向及其变化对于渔船不同行为的轨迹段提取具有重要作用^[17]。渔船进出港轨迹行为的主要特征为:①在航速方面,由于靠近渔海岸线但仍处于航行状态,因此航速相较于停泊行为略高,远小于往返于渔场之间和作业前后的航速,且通常处于加速或减速状态;②在航首向方面,因为渔船在渔港附近,船头一般为面向或背离渔海岸线的方向,且航行轨迹多呈直线,不会发生频繁的转向,因此航首向通常集中分布于一个较大或较小的范围内。

3 渔船进出港区域识别方法

该文整体研究流程如图 2 所示。

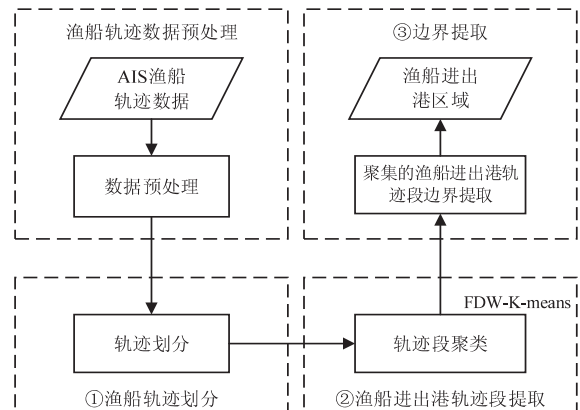


图 2 渔船进出渔港区域识别整体流程

包括:①基于多特征融合下轨迹点间相似性的轨

迹划分算法,将轨迹划分为具有不同渔船行为表现的轨迹段,不同渔船行为包括停泊行为、作业行为和航行行为等;②提出 FDW-K-means 聚类算法,将上一步得到的轨迹段的特征作为聚类对象,实现渔船进出港轨迹段的提取;③综合运用 DBSCAN 聚类算法、Del-Alpha-Shape 算法提取聚集的渔船进出港轨迹段的边界,从而获得渔船进出港区域。

3.1 基于多特征融合下轨迹点间相似性的轨迹行为划分算法

算法思想:计算渔船轨迹每两个轨迹点之间的时间距离、空间距离、航速距离和航首向距离等 4 个特征距离,通过加权求和的方式将特征距离融合起来作为轨迹点之间的特征相似距离;然后基于核心距离的思想,通过计算各轨迹点核心距离判断轨迹点之间的相似性,当两点核心距离差值的绝对值大于阈值时,说明两个轨迹点相似程度低,即表示两个轨迹点行为不同,从而达到对渔船轨迹划分的目的。划分后的按采集时间排序的轨迹点组成的轨迹段为一个渔船行为轨迹段,连续的轨迹段具有不同的渔船行为。

由第 2.2 节可知,航速和航首向可以作为划分渔船轨迹不同行为的典型特征。同时,考虑到渔船轨迹具有时空属性,因此选择时间、空间、航速和航首向等 4 个典型属性进行研究。算法实现的步骤如下。

该算法输入为各渔船轨迹。其中轨迹点的形式为 (shipname, utc, lon, lat, v, h), shipname, utc, lon, lat, v, h 分别表示渔船编号、采集时间戳、经度、纬度、航速和航首向。输出为将渔船轨迹划分得到的不同渔船行为轨迹段。

第 1 步 遍历当前航次数据,根据公式 1~4 分别计算轨迹点与其他各轨迹点之间的时间距离 $\text{dis}_T(P_i, P_j)$ 、空间距离 $\text{dis}_S(P_i, P_j)$ 、航速距离 $\text{dis}_V(P_i, P_j)$ 和航首向距离 $\text{dis}_H(P_i, P_j)$ 。每个轨迹点以 (utc_i, lon_i, lat_i, $\text{dis}_T(P_i, P_j)$, $\text{dis}_S(P_i, P_j)$, $\text{dis}_V(P_i, P_j)$, $\text{dis}_H(P_i, P_j)$) 保存。

$$\text{dis}_T(P_i, P_j) = |\text{utc}_i - \text{utc}_j| \quad (1)$$

$$\text{dis}_S(P_i, P_j) = R \times \arccos[\cos \text{lat}_i \cos \text{lat}_j \cos(\text{lon}_i - \text{lon}_j) + \sin \text{lat}_i \sin \text{lat}_j] \quad (2)$$

$$\text{dis}_V(P_i, P_j) = |v_i - v_j| \quad (3)$$

$$\Delta h = |h_i - h_j|$$

$$\text{dis}_H(P_i, P_j) = \begin{cases} \Delta h, & \Delta h \leq 180 \\ 360 - \Delta h, & \Delta h > 180 \end{cases} \quad (4)$$

其中,空间距离 $\text{dis}_S(P_i, P_j)$ 采用球面两轨迹点间距离计算公式, R 表示地球半径,大小为 6 371.393 km^[18],航速距离 $\text{dis}_V(P_i, P_j)$ 采用速度差值的绝对值求得^[19]。 P_i, P_j 分别表示渔船轨迹中第 i, j 个轨迹点。

第 2 步 根据公式 5 计算每个轨迹点分别与其他轨迹点之间的特征相似距离 d_{ij} 。

$$d_{ij} = \theta_T \times \text{dis}_T(P_i, P_j) + \theta_S \times \text{dis}_S(P_i, P_j) + \theta_V \times \text{dis}_V(P_i, P_j) + \theta_H \times \text{dis}_H(P_i, P_j) \quad (5)$$

$$\theta_T + \theta_S + \theta_V + \theta_H = 1 \quad (6)$$

其中, d_{ij} 表示 P_i, P_j 两个轨迹点之间的特征相似距离。特征相似距离实现了轨迹点之间多特征的融合,为下一步判断轨迹点之间的相似性提供数据支持。 θ_T 、 θ_S 、 θ_V 、 θ_H 分别表示时间距离、空间距离、航速距离和航首向距离的权重,表示各特征距离对特征相似距离的影响比重,并满足公式 6。权重值的大小取决于轨迹行为对特征距离的依赖程度,因为航速和航首向对轨迹行为的影响较大,所以适当增大航速距离权重 θ_V 和航首向距离权重 θ_H 。

第 3 步 基于核心距离思想^[20] 计算各轨迹点的核心距离 d_{icount_n} 。在该文中计算方法是令轨迹点为中心点,将轨迹点的特征相似距离 d_{ij} 进行增序排列后,该点到满足邻域点数阈值 count_n 的另一个轨迹点之间的距离。核心距离的大小决定了该点位置的空间密度,相邻轨迹点的空间密度越接近,表示轨迹点之间相似程度越高。在这里点数阈值 count_n 满足公式 7, n 为该渔船轨迹的轨迹点数目。

$$\text{count}_n = 5\%n \quad (7)$$

第 4 步 根据公式 8 计算划分阈值 disSeg,当轨迹点 $|d_{\text{icount}_n} - d_{(i+1)\text{count}_n}| \geq \text{disSeg}$ 时,在点 P_i 和点 P_j 之间对轨迹进行划分,两个划分处之间按采集时间排序的连续轨迹点组成一个行为的轨迹段。该渔船轨迹遍历完成时,执行第 1 步,直到遍历完所有渔船轨迹。

$$\text{disSeg} = \frac{1}{n-1} \sum |d_{\text{icount}_n} - d_{(i+1)\text{count}_n}| \times k \quad (8)$$

其中,划分阈值为渔船轨迹各相邻轨迹点之间核心距离差值的平均值的 k 倍, k 取值一般为 10^[4]。

基于多特征融合下轨迹点间相似性的轨迹划分算法流程如图 3 所示。

3.2 FDW-K-means 聚类算法

K-means 算法通过重定位聚类簇中心的方式,将聚类对象划分到各个簇,使得聚类对象到各簇中心距离最小,从而达到相同类型对象聚合的目的^[12]。因为由 3.1 节得到的一个轨迹段的轨迹点具有相同渔船行为,因此各轨迹点在特征变化上具有同步性,为避免波动数据点对结果的影响,将划分后的各轨迹段的特征作为聚类对象。根据不同特征对聚类结果影响不同,该文提出 FDW-K-means 聚类算法,从而实现渔船进出港轨迹段的提取。算法思想是在计算轨迹段与每个聚类簇中心各特征的距离时,采用增加权重的方式进行重要性判断,通过权重衡量各特征对聚类结果的影

响,从而提高聚类结果的准确程度。权重值作为参数参与聚类过程。

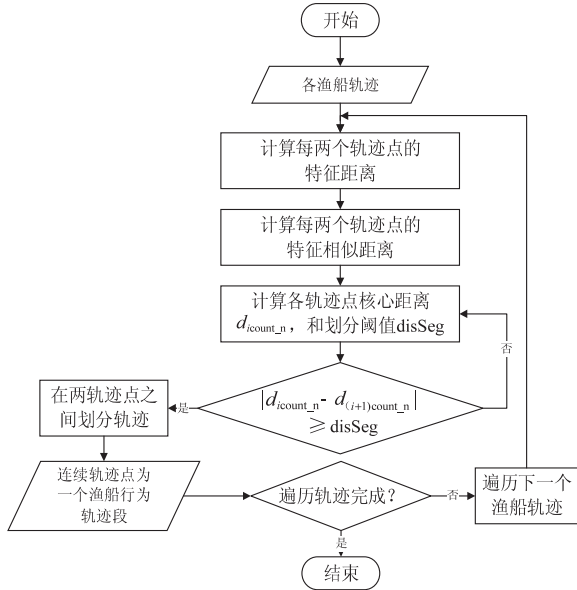


图 3 轨迹划分算法流程

需要指出的是,该阶段主要通过 FDW-K-means 实现渔船进出港轨迹段的提取,无需对研究对象进行更多其他行为的聚类,只需实现渔船进出港轨迹段和其他行为的轨迹段的聚类。因此设置聚类个数 K 为 2。根据聚类后各簇的特征表现将结果分为渔船进出港轨迹段和非进出港轨迹段。输出结果为渔船进出港轨迹段轨迹点集 TG_{cl} 和其他非进出港轨迹段轨迹点集 TG_{other} 。

根据 2.2 节渔船航行特征分析,采用各轨迹段航速、航首向平均值作为轨迹段特征。

假设待聚类轨迹段集合为 $\{T_1, T_2, \dots, T_l\}$, 其中 l 表示待聚类轨迹段的个数。轨迹段特征为航速平均值 \bar{v} 和航首向平均值 \bar{h} 。初始化的聚类中心为 $\{C_1, C_2\}$, $C_t = \{C_{t_v}, C_{t_h}\}$ ($t = 1, 2$)。已知各轨迹段有两个属性维度,则每个轨迹段到每个聚类中心特征属性加权后的距离计算方法如公式 9 所示:

$$\text{dis}(T_i, C_t) = \sqrt{w_v (\bar{v}_i - C_{t_v})^2 + w_h (\bar{h}_i - C_{t_h})^2} \quad (9)$$

其中, $i = 1, 2, \dots, l$, l 为轨迹段的个数, C_{t_v} 和 C_{t_h} 分别表示作为初始聚类中心的轨迹段的平均航速和平均航首向, w_v 和 w_h 分别表示航速距离和航首向距离的权重,且 w_v, w_h 取值范围为 $[0, 1]$ 。权重变化条件为 w_value , 当航速权重增加 w_value 时,航首向权重减小 w_value , 并满足公式 10。

$$w_v + w_h = 1 \quad (10)$$

每一次聚类后,各聚类中心各个特征计算方法如公式 11 和公式 12 所示:

$$C_{t_v} = \frac{\sum_{v_i \in S_t} \bar{v}_i}{|S_t|} \quad (11)$$

$$C_{t_h} = \frac{\sum_{h_i \in S_t} \bar{h}_i}{|S_t|} \quad (12)$$

FDW-K-means 聚类算法流程如图 4 所示。

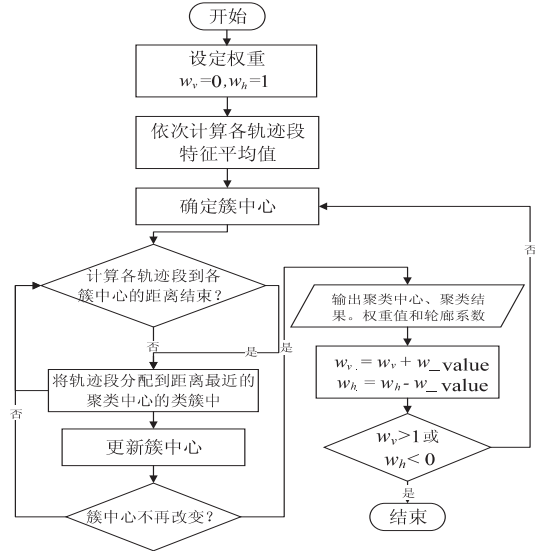


图 4 FDW-K-means 算法流程

3.3 渔船进出渔港区域识别方法

算法思想:聚类后得到的渔船进出港轨迹段数据分布在渔港海岸线附近。对同一渔船进出港区域而言,不同渔船或同一渔船在该区域产生的渔船进出港轨迹段位置和数目也不相同,不同位置的渔船进出港轨迹段可能属于同一个渔船进出港区域,所以需要对 3.2 节得到的渔船进出港轨迹段进行聚类处理。对聚类后的每一个簇进行边界提取和中心点提取,边界所围区域为一个渔船进出港区域,中心点的经纬度数据表示渔船进出港区域的位置。

该文采用 DBSCAN 算法对 3.2 节所得渔船进出港轨迹段轨迹点集 TG_{cl} 进行聚类处理,从而生成各位置的渔船进出港轨迹段聚类簇;采用文献[5]提出的 Delauney 三角剖分算法与 Alpha Shapes 轮廓线提取算法相结合的 Del-Alpha-Shape 算法对各聚类簇进行边界提取从而获得各渔船进出渔港区域。渔船进出港区域识别算法具体流程为:

输入:渔船进出港轨迹段轨迹点集 TG_{cl} ;

输出:各渔船进出港轨迹段簇边界轨迹点集 $TG_{bound} = \{P_1, P_2, \dots, P_x\}$, 其中 P 为边界轨迹点,包括经纬度、时间戳等信息, x 为边界点个数。各渔船进出港轨迹段簇中心点 P_c , 包括经纬度信息。

第 1 步:基于 DBSCAN 聚类算法对渔船进出港轨迹段轨迹点集 TG_{cl} 进行聚类处理。

第 2 步:遍历每一个渔船进出港轨迹段簇。设定

任意两点外接圆半径 α 值,采用 Delauney 三角剖分算法对簇数据进行三角剖分。根据公式 13 计算各三角形密度指数 ρ ,根据公式 14 计算密度指数阈值为 ρ_z ,满足 $\rho \geq \rho_z$ 时说明该三角形是渔船进出港区域三角形,保留三角形端点到 listTri。然后根据公式 15 和公式 16,计算 listTri 内轨迹点的经纬度极值的平均值,得到各渔船进出港轨迹段簇的中心点的经纬度^[21]。

$$\rho = \frac{1}{\text{circum_r}} \quad (13)$$

$$\rho_z = \frac{1}{\alpha} \quad (14)$$

$$\text{lon}_{\text{center}} = \frac{\text{lon}_{\text{max}} + \text{lon}_{\text{min}}}{2} \quad (15)$$

$$\text{lat}_{\text{center}} = \frac{\text{lat}_{\text{max}} + \text{lat}_{\text{min}}}{2} \quad (16)$$

其中,circum_r 为三角形外接圆半径;lon_{center} 和 lat_{center} 分别为簇中心点的经度和纬度;lon_{max} 和 lon_{min} 分别为簇内经度的极大值和极小值;lat_{max} 和 lat_{min} 分别为簇内纬度的极大值和极小值。

第 3 步:采用 Alpha Shapes 算法对第 2 步所得 listTri 中轨迹点进行边界提取。重复第 2 步,直至各簇数据遍历完成。输出的是各簇边界轨迹点集,相邻边界轨迹点连线所围区域即为渔船进出港区域。同时输出各簇中心点。

渔船进出港区域识别算法流程如图 5 所示。

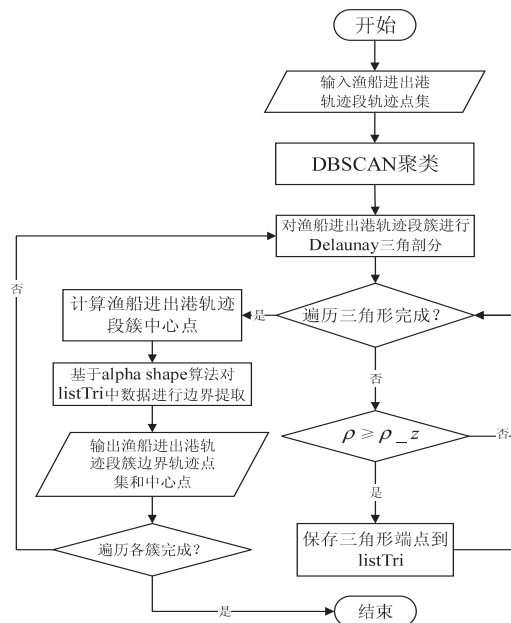


图 5 渔船进出港区域识别算法流程

4 实验

4.1 实验环境和数据集

实验环境是 Windows 10 操作系统,CPU 为 Intel

(R) Core(TM) i7-1165G7,主频为 2.80 GHz,内存大小为 16 G,以 python3.6 编程语言实现。实验使用的是 2021 年 3 月份地理区域为椒江渔港、博贺渔港的包括 1 035 艘渔船的 AIS 渔船轨迹数据,每个文件包含了 3 月份各渔船全天采集时刻的轨迹数据信息,总共约 3 000 000 余条记录。

实验对采集到的原始渔船轨迹数据进行预处理,删除噪声点,对缺失的数据进行插值补全,并提取有用字段。经过预处理后,每条记录处理为表 1 中 6 个字段。因为渔船进出港区域位于渔港附近且渔船处于航行状态,为了提高方法效率,按照渔船轨迹点到港口的距离和航速进行了筛选,选取渔港周围 10 海里以内,不包括较远海域的轨迹点和处于高速航行(航速大于 8 节)的轨迹点。因此经过渔船轨迹划分实验后,采用的椒江渔港轨迹段 21 768 个,博贺渔港轨迹段 23 179 个。

表 1 数据预处理后的数据基本信息

名称	含义	举例	字段说明
shipname	船名	浙椒渔 76009	船名
utc	时间戳	1 614 745 811	采集时间
lon	经度	121.453 621	渔船采集时刻所在经度
lat	纬度	28.688 615	渔船采集时刻所在纬度
v	航速	9.6	对地航速
h	航首向	135	对地航向

4.2 评价标准

4.2.1 FDW-K-means 算法和渔船进出港轨迹段提取结果评价标准

(1)将传统 K-means 算法作为 FDW-K-means 算法的对比方法,采用所有聚类对象轮廓系数平均值来评价 FDW-K-means 聚类算法的聚类效果,通过轮廓系数的对比来验证该文所提 FDW-K-means 算法的有效性。轮廓系数是指经过聚类后不同类型对象的稠密程度和离散程度^[22]。计算方法如公式 17:

$$s = \frac{b - a}{\max(a, b)} = \begin{cases} 1 - a/b, & a < b \\ 0, & a = b \\ b/a - 1, & a > b \end{cases} \quad (17)$$

其中, s 表示轮廓系数, a 表示聚类后该轨迹段与聚类到同一个簇的轨迹段之间的特征距离平均值, b 表示聚类后该轨迹段与聚类到另一个簇的轨迹段之间的特征距离平均值。轮廓系数的取值范围是 $(-1, 1)$,轮廓系数越接近于 1,表示聚类效果越好。

(2)通过对聚类后得到的渔船进出港轨迹段特征表现来评价渔船进出港轨迹段提取结果。

4.2.2 渔船进出港区域识别方法评价标准

采用以下两种渔船进出港区域识别方法作为该文

所提方法的对比方法,采用数据集不变。

(1)基于传统 K-means 算法提取渔船进出港轨迹段从而实现渔船进出港区域识别的方法;

(2)文献[9]所提的基于对各轨迹点特征设定约束条件的思想实现渔船进出港区域识别的方法。

将采用该文方法和对比方法识别到的椒江渔港、博贺渔港渔船进出港区域,叠加到当前文献[1]中利用人工或遥感影像等方式实际划定并采用的渔船进出港区域的影像图上检测其正确性,并通过识别到的各渔港的渔船进出港区域位置正确时数量的正确率来表示方法的有效性。其中,实际划定并采用的渔船进出港区域位置为测试集,用于检测渔船进出港区域的识别结果;正确率计算方法为采用该文方法识别到的各渔港的渔船进出港区域位置与当前实际划定并采用的渔船进出港区域位置相同的数量占识别到的渔船进出港区域数量的百分比。正确率越高表示方法越有效。

4.3 参数设置

实验构建 FDW-K-means 算法来实现轨迹段的聚类。K 值设置为 2,聚类结果一类为渔船进出港轨迹段,另一类为非进出港轨迹段;由于 K 值为 2,所以初始化中心点运算次数为默认值 10;在实验过程中对聚类效果评价,聚类中心稳定时停止,并得到权重参数。

4.4 实验结果与分析

4.4.1 FDW-K-means 聚类轮廓系数结果

图 6 为基于 FDW-K-means 算法对不同行为轨迹段进行聚类时得到的轮廓系数变化图。以航速距离权重值变化为主进行统计,实验设置取值范围为[0,1],w_value 为 0.05。从图 6 可以得出:①当航速距离权重值较大时,轮廓系数越接近于 1,说明聚类效果越好;②当航速距离权重值为 0.65 时,轮廓系数最高,为 0.952 3,此时对应的航首向距离权重值为 0.35;③航速特征相较于航首向特征对聚类效果的影响更重要。

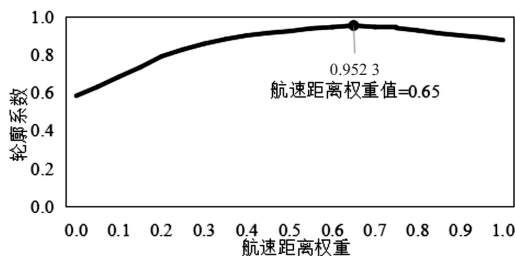


图 6 不同权重的聚类轮廓系数

将 FDW-K-means 轮廓系数与采用传统 K-means 算法聚类所得轮廓系数进行对比,结果如表 2 所示。根据对比结果可知,FDW-K-means 聚类结果轮廓系数相较于传统 K-means 算法提高了将近 0.25,表示将渔船轨迹段的不同特征通过增加权重的形式进行重要性判断,可以考虑到各个特征的影响程度,会达到更好

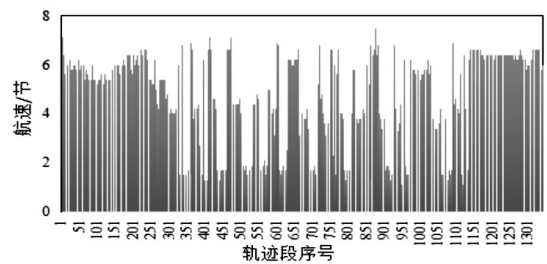
的聚类效果。

表 2 不同模型的聚类轮廓系数对比

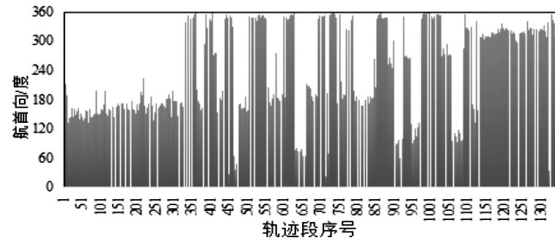
模型	轮廓系数
K-means	0.713 6
FDW-K-means	0.952 3

4.4.2 渔船进出港轨迹段提取结果

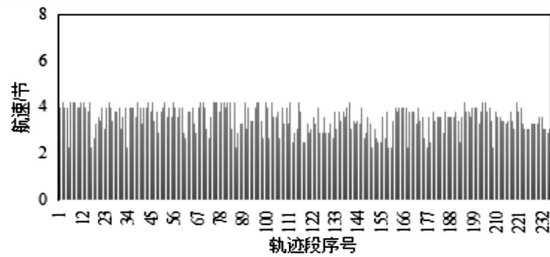
图 7 所示为部分轨迹段聚类后效果展示柱状图,其中横坐标是轨迹段序号,纵坐标分别是每个轨迹段的平均航速和平均航首向。图 7(a)和图 7(b)分别为待聚类轨迹段的平均航速和平均航首向图。图 7(c)和图 7(d)分别为聚类后得到的渔船进出港轨迹段的平均航速和平均航首向图。



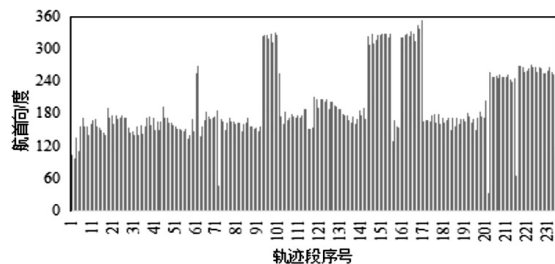
(a) 轨迹段平均航速



(b) 轨迹段平均航首向



(c) 进出港行为轨迹段平均航速



(d) 进出港行为轨迹段平均航首向

图 7 轨迹段聚类结果

从图 7 中可以得出:①从平均航速来看,渔船进出港轨迹段主要集中在 2~4 节之间,相较于停泊行为等其他行为的轨迹段区别明显;②从平均航首向来看,渔

船进出港轨迹段的航首向大小分布更为集中,表明大部分渔船在进港或出港时主要集中在某个方向;③从平均航速和平均航首向对应关系来看,可以根据航首向判断渔船处于的状态是进港或出港,当渔船航首向偏大时,一般处于进港状态,当航首向偏小时,一般处于出港状态。上述三点结论与 2.2 节所提到的渔船进出港行为表现和文献[4]中有关渔船进出港行为特征表现的介绍基本一致,证明了该文渔船进出港轨迹段提取的有效性。

4.4.3 渔船进出港区域识别结果

根据实验调整 α 取值,当 $\alpha = 210$ 时,将识别到的椒江渔港和博贺渔港的渔船进出港区域位置与采用文献[1]方法采集的实际渔船进出港区域位置进行对比,得到的位置正确数量的正确率最好,分别为 94.2% 和 95.8%。图 8 为识别到的椒江渔港和博贺渔港的某渔船进出港区域图。

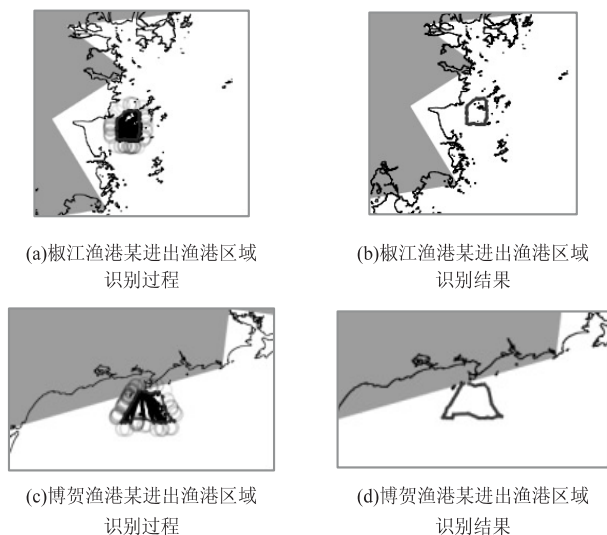


图 8 渔船进出渔港区域识别结果

其中,图 8(a) 和图 8(c) 分别为对渔船进出港轨迹段簇边界的提取过程,图 8(b) 和图 8(d) 分别为识别到的渔船进出港区域。

表 3 为采用该文研究方法和对比方法识别到的椒江渔港和博贺渔港的渔船进出港区域的正确率。

表 3 使用不同方法判断渔船进出港情况的正确率

渔港名称	使用方法	正确率/%
椒江渔港	该文研究方法	94.2
	传统 K-means	83.5
	文献[9]方法	85.5
博贺渔港	该文研究方法	95.8
	传统 K-means	86.3
	文献[9]方法	89.2

从表 3 可以得出:①通过与应用传统 K-means 算法提取渔船进出港轨迹段从而识别渔船进出港区域的

方法对比来看,应用该文所提方法识别到的椒江渔港和博贺渔港的渔船进出港区域正确率分别提高了 10.7% 和 9.5%,说明该文提出的 FDW-K-means 算法通过对特征增加权重的方式,能够更好地判断各特征对轨迹段聚合的影响,实现了有效的渔船进出港轨迹段提取,从而使得识别到渔船进出港区域的正确率更高;②通过与使用文献[9]基于对各轨迹点特征设定约束条件思想识别渔船进出港区域的方法对比来看,该文研究方法正确率分别提高了 8.7% 和 6.6%,说明将各轨迹点按离散且孤立的点来处理,只考虑其各自特征来分析轨迹行为具有一定的局限性,还需要考虑到轨迹点之间的时空关系,才能达到更好的渔船进出港区域的识别效果;③从表 3 结果来看,该文研究方法识别到的椒江渔港和博贺渔港的渔船进出港区域的正确率分别为 94.2% 和 95.8%,说明采用该文研究方法可以实现渔船进出港区域的有效识别。

综上,该文提出的渔船进出港区域识别方法能够考虑到轨迹点之间的时空关系和轨迹段不同特征的影响程度,实现有效的渔船进出港区域识别。

5 结束语

该文提出了一种基于渔船轨迹数据的渔船进出港区域识别方法。提出的基于多特征融合下轨迹点间相似性的轨迹划分算法和 FDW-K-means 聚类算法相结合提取渔船进出港轨迹段,综合运用 DBSCAN 聚类算法和 Del-Alpha-Shape 算法最终得到渔船进出港区域。相较于传统基于轨迹数据进行区域识别的方法,该文所提方法能够更好地利用轨迹点之间的时空关系,还能考虑到轨迹段不同特征的影响程度,使得识别到的渔船进出港区域正确率更高。理论分析与基于真实数据集的实验结果证明了该方法的有效性。能够为渔船进出港监管提供有效的信息支撑和科学的数据参考。但是该文所提方法还存在部分渔船进出港区域提取位置不正确的情况。在下一步的工作中,将以不同大小和作业类型的渔船轨迹数据为基础分类开展研究,挖掘不同类型渔船的进出港轨迹段,实现更精确的渔船进出港区域识别,提高该方法的适用性。

参考文献:

- [1] 吴坤阳. 虚拟“电子围栏”在船舶进出港报告监管中的应用[J]. 交通企业管理, 2020, 35(2): 98-100.
- [2] 肖 潇, 邵哲平, 潘家财, 等. 基于 AIS 信息的船舶轨迹聚类模型及应用[J]. 中国航海, 2015, 38(2): 82-86.
- [3] 张 圣, 王列辉. 船舶自动识别系统数据在港口地理学研究中的应用与展望[J]. 大连海事大学学报, 2023, 49(3): 61-73.
- [4] 吴宝福. 基于轨迹数据的渔船行为判别关键技术研究

- [D]. 杭州:杭州电子科技大学,2019.
- [5] 徐 垚,李卓然,孟金龙,等. 基于大规模船舶轨迹数据的航道边界提取方法[J]. 计算机应用,2019,39(1):105-112.
- [6] ALVARES L O,BOGORNY V,KUIJPERS B,et al. A model for enriching trajectories with semantic geographical information[C]//Proceedings of the 15th annual ACM international symposium on advances in geographic information systems. Seattle:ACM,2007:1-8.
- [7] ZHENG Y,ZHANG L,XIE X,et al. Mining correlation between locations using human location history[C]//Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems. Seattle:ACM,2009:472-475.
- [8] ROCHA J,TIMES V C,OLIVEIRA G,et al. DB-SMoT: a direction-based spatio-temporal clustering method[C]//5th IEEE international conference on intelligent systems. London:IEEE,2010.
- [9] 郑振涛,赵卓峰,王桂玲,等. 面向港口停留区域识别的船舶停留轨迹提取方法[J]. 计算机应用,2019,39(1):113-117.
- [10] PALMA A T,BOGORNY V,KUIJPERS B,et al. A clustering based approach for discovering interesting places in trajectories[C]//Proceedings of the 2008 ACM symposium on applied computing. Fortaleza:ACM,2008:863-868.
- [11] 张治华. 基于 GPS 轨迹的出行信息提取研究[D]. 上海:华东师范大学,2010.
- [12] 周志华. 机器学习[M]. 北京:清华大学出版社,2016:197-205.
- [13] 耿家利. 渔船轨迹大数据存储优化与行为识别技术研究[D]. 杭州:杭州电子科技大学,2018.
- [14] 王代楠,陈 琼. 改进 K-means 算法的船舶密集区域识别模型研究与 VTS 应用[J]. 中国海事,2022(1):53-56.
- [15] 杨 伟,艾廷华. 众源车辆轨迹加油停留行为探测与加油站提取[J]. 测绘学报,2017,46(7):918-927.
- [16] 惠振阳,胡海瑛,李 娜,等. 改进的 Alpha-shapes 建筑物轮廓线提取算法[J]. 激光与光电子学进展,2022,59(10):447-455.
- [17] 崔彤彤,王桂玲,高 晶. 基于 1DCNN-LSTM 的船舶轨迹分类方法[J]. 计算机科学,2020,47(9):175-184.
- [18] 汤先峰,张胜茂,樊 伟,等. 基于北斗船位数据的江苏渔船航次分析[J]. 渔业现代化,2020,47(1):63-71.
- [19] 纪晓妍. 基于 CBR 的 AIS 船舶到港时间预测模型研究[D]. 大连:大连海事大学,2022.
- [20] ANKERST M,BREUNIG M M,KRIEGEL H P,et al. OPTICS: ordering points to identify the clustering structure [C]//Proceedings ACM SIGMOD international conference on management of data. Philadelphia:ACM,1999.
- [21] 刘鑫鑫,韩 懿. 基于 DBSCAN 的港口泊位自动识别算法设计[J]. 上海船舶运输科学研究所学报,2021,44(1):31-36.
- [22] 孙 林,刘梦含,徐久成. 基于优化初始聚类中心和轮廓系数的 K-means 聚类算法[J]. 模糊系统与数学,2022,36(1):47-65.