

阈值引导采样法的船舶轨迹简化算法

张银昊, 潘家财, 赵梦鸽

(集美大学航海学院, 福建 厦门 361021)

[摘要] 为了解决大多数的轨迹简化算法缺乏对轨迹多个特征的综合衡量的问题, 提出了一种融合位置、方向、速度与时序等多个轨迹特征, 基于阈值引导采样的船舶轨迹简化算法。该算法以角度阈值法思想为基础, 计算轨迹点的前后多个特征差, 与设置的阈值进行对比, 进而实现对轨迹的简化。利用船舶AIS轨迹数据对该算法进行轨迹简化和轨迹聚类实验。实验结果表明, 该轨迹简化算法在简化率、简化误差率上均优于角度阈值算法, 能保留原始轨迹的形状, 并且简化后的轨迹数据能运用于轨迹聚类, 聚类效果良好。

[关键词] 船舶轨迹; 简化算法; 阈值引导采样法; 船舶自动识别系统; 移动对象

[中图分类号] U 675.7

A Ship Trajectory Simplified Algorithm Based on Threshold Guiding Sampling Method

ZHANG Yin hao, PAN Jia cai, ZHAO Meng ge

(Navigation College, Jimei University, Xiamen 361021, China)

Abstract: In order to solve the lack of comprehensive measurement of multiple features of trajectory in most simplified algorithms, we propose a trajectory simplification algorithm which is based on the angle threshold algorithm and takes into account multiple trajectory features such as position, direction, speed and time sequence. The proposed algorithm calculates the differences between features of the successive trajectory points and compares them with the set thresholds to simplify the trajectories. Using ship's AIS trajectory data, trajectory simplification and trajectory clustering are performed on the proposed algorithm. The results of the experiments show that the proposed algorithm is superior to the angle threshold algorithm in terms of simplification rate and simplification error rate. The results also show that it can retain the shape of the original trajectory and the simplified trajectory can effectively be used for trajectory clustering.

Keywords: ship trajectory; simplification algorithm; threshold guide sampling algorithm; ship automatic identification system; moving object

0 引言

船舶AIS(automatic identification system)数据蕴含丰富的船舶信息, 利用这些数据能够发现和分析船舶的行为, 从而为海事部门对船舶航行的安全监管和决策提供支持服务^[1-2]。船舶AIS数据具有

[收稿日期] 2021-02-09

[作者简介] 张银昊(1995—), 男, 硕士生, 从事船舶轨迹数据挖掘研究。通信作者: 潘家财(1975—), 男, 教授, 硕士, 从事航海技术和海上交通信息研究。E-mail: panjiacai@163.com。

量大、实时、多样等特征,准确地存储和提取这些数据是分析船舶轨迹的前提,但是海量的数据会使查询和计算效率降低,因此,通常需要采用轨迹简化的方法来处理轨迹数据,进而保留关键数据点,减少轨迹数据存储的负担。目前,国内外学者对船舶轨迹简化进行了许多研究。文献 [3] 提出了使用垂直欧式距离 (perpendicular euclidean distance) 进行轨迹简化;时间同步欧式距离 (time synchronized euclidian distance)^[4] 是基于时间同步的轨迹计算方式来产生近似轨迹。Zhao^[5] 基于 douglas-peucker 模型^[6] 建立轨迹点批处理模式,但在实际的轨迹简化处理中,很难实现批处理模式,该研究仅仅衡量了船舶的经纬度的位置信息,忽视了航速和航向信息。潘家财^[7] 根据船舶航行的特征,对航向和航速变化率设置阈值进行特征点的选取;而肖潇^[8] 利用最小描述长度准则^[9] (minimum description length, MDL) 进一步在航速、航向变化率之上筛选特征点,但是该方法缺少对 AIS 信息多因素的综合考虑。角度阈值法^[10] 是一种以轨迹点的角度变化作为衡量标准的轨迹简化方法,计算每一个轨迹点的角度变化量,若大于阈值,当前轨迹点视为特征点。

轨迹简化主要运用于轨迹聚类,轨迹聚类主要分为两种:1) 将船舶的轨迹视为整体进行聚类^[11];2) 将船舶轨迹进行分段,分别对分段后的轨迹子段进行聚类分析,将相似的轨迹子段归类为簇,保证轨迹运动的细节信息不丢失^[12-15]。目前,对两种简化方式均能兼顾的轨迹简化方法较少。

本文根据船舶移动轨迹的运动特征,在角度阈值法^[10] 的基础上提出一种新的轨迹简化方法——阈值引导采样法,并采用聚类实验方法进行验证。实验结果表明,该方法在轨迹段聚类和整体轨迹聚类的应用上均有较好的聚类效果。

1 阈值引导采样算法

1.1 相关理论

阈值引导采样法是在角度阈值法^[10] 的基础上提出的一种新的简化方法。角度阈值法^[10] 通过设置角度的阈值,并将当前点与上一个点之间的角度进行对比,角度差大于阈值时,当前轨迹点视为特征点。阈值引导采样法沿用了角度特征判断的特点,设置了船舶转向角变化的角度阈值 θ_{\max} , 进行角度特征的判定;同时,还增设了速度阈值 S_{\max} , 以及方位角阈值 β_{\max} , 从船舶速度及转角两个方面对船舶轨道特征点做进一步判断。

船舶数据出现时间位置不合理的情况主要为:在较大时间区间内反复进出研究水域的边界;在较大时间段内突然丢失船舶的轨迹数据;仅有一个数据点的船舶。为解决这个问题,设置了中断特征点 (Interrupted feature point), 为了判断中断特征点,对轨迹设置了时间间隔阈值 t_{\max} , 准确地分割轨迹。

1.2 参数定义

1.2.1 初始轨迹信息

原始轨迹 $T_i = \{p_1, \dots, p_j, \dots, p_n\}$, $p_j = (i, s_j, c_j, t_j, \text{lon}_j, \text{lat}_j)$, 其中: T_i 是船舶为 i 的轨迹数据点集合; p_j 为轨迹 i 的第 j 个轨迹点; s_j 为轨迹点 j 的 SOG (speed over ground); c_j 为轨迹点 j 的 COG (course over ground); t_j 为轨迹点 j 的时刻; lon_j 和 lat_j 分别表示轨迹点 j 的经度和纬度。

简化后的轨迹信息: $T_{m_{\text{simp}}} = \{p_{m_{\text{simp}}_1}, p_{m_{\text{simp}}_2}, \dots, p_{m_{\text{simp}}_k}, \dots, p_{m_{\text{simp}}_d}\}$, $p_{m_{\text{simp}}_k}$ 表示 T_i 中的特征点。

1.2.2 角度阈值和速度阈值

对于 T_i 中的轨迹点 p_j , 满足 $C_{\text{diff}} \geq \theta_{\max}$, 则 p_j 视为角度特征点。其中: 角度差 $C_{\text{diff}} = |c_{j-1} - c_j|$; θ_{\max} 为角度阈值。

对于 T_i 中的轨迹点 p_j , 满足 $S_{\text{diff}} \geq S_{\max}$, 则 p_j 视为速度特征点。其中: 速度差 $S_{\text{diff}} = |s_{j-1} - s_j|$; S_{\max} 为速度阈值。

通常角度阈值和速度阈值需要根据研究对象的运动习惯来确定。例如,船舶的极限舵角通常为 $30^\circ \sim 35^\circ$, 航行时船舶改变航向次数往往多于改变航速次数。因此,角度阈值一般设置为 $0^\circ \sim 5^\circ$, 若为了追求更简洁的简化轨迹,可将角度阈值设为 15° , 而速度阈值一般在 $0 \sim 3 \text{ kn}$ 内选取。

1.2.3 中断特征点

对于 T_i 中的轨迹点 p_j , 满足 $t_{\text{diff}} \geq t_{\text{max}}$, 则 p_j 视为中断特征点, $p_j = p_{m_{\text{simp_d}}}$, $p_{j+1} = p_{m+1_{\text{simp_l}}}$ 。其中: 时间差 $t_{\text{diff}} = |t_{j-1} - t_j|$; t_{max} 为时间间隔阈值。选取 A 级船长 230 m 以上船舶的 AIS 数据, 该类船舶自主模式下信息更新率最大为 180 s^[16]。中断特征点以时间差为阈值进行判断, 再考虑误差, t_{max} 一般选取为 180~300 s。

1.2.4 方位角阈值

对于 T_i 中的轨迹点 p_j , 满足 $\beta_{\text{diff}} \geq \beta_{\text{max}}$, 则 p_j 视为方位角特征点。其中: β_{max} 为方位角阈值, 轨迹线段的坐标方位角差为 $\text{azi}_{ij} = |c_i - c_j|$, 方位角差 $\beta_{\text{diff}} = |\text{azi}_{i(i+1)} - \text{azi}_{ij}|$, c_i 为 p_j 的上一个特征点的 COG, c_j 为 p_j 的 COG。通常方位角阈值的大小需要根据研究水域的转向点的角度来确定, 例如, 当转向点处两航道的夹角为 20° 时, 为了保留一定的空间, 方位角阈值设置需要在夹角基础上增加约 10°, 即方位角阈值设为 30°, 然后在这个阈值附近进行调整, 进行多次的实验, 选取更精确的阈值。

1.3 轨迹点特征优先级的判断

在轨迹简化过程中, 为避免数据点出现的时间位置不合理而被视为一条轨迹的情况, 应优先判断数据点是否为中断特征点, 将轨迹进行分割。在船舶轨迹分析中, 船位的突变必然导致航速或航向的异常, 但船速和航向的突变未必会带来船位的异常。根据信息熵理论^[17], 船位异常事件所携带的信息大于航速与航向异常事件的信息量, 在判断优先级上优先判断信息量较大的属性, 能减少运算量, 提高运算效率。因此, 需要优先判断数据点的方位角特征, 再判断角度与速度特征。

轨迹点特征判断具体如下:

1) 判断点的方向和速度位于上一个点的方向和速度的阈值区间内, 则该点视为可简化。如图 1 所示, O 、 A 、 B 、 C 、 D 为一条船舶的 5 个 AIS 轨迹点, 按照时序排列; V_{min} 为上一个点的 SOG; $V_{\text{max}} = V_{\text{min}} + S_{\text{max}}$; $2\theta_{\text{max}}$ 的中分线方向为上一个点的 COG; F_x 表示船舶在点 X 的 COG。由于 O 、 D 两点为起点与终点, 无需判断, 均视为特征点。从第二个点 A 开始判断是否属于特征点, F_A 的长度为船舶在 A 点的 SOG, 当 F 的长度没有落在 $(V_{\text{min}}, V_{\text{max}})$ 内时, A 点因速度变化过大而视为轨迹的特征点; 同理, F_A 的方向没有落在 $2\theta_{\text{max}}$ 范围内, 则 A 点因角度变化过大而视为轨迹的特征点。因此, 当判断点不满足以上任意一个条件, 就将该点视为轨迹特征点。同理, 下一个待判断点 B , 将上一个点 A 的 SOG 视为 V_{min} , 点 A 的 COG 视为 $2\theta_{\text{max}}$ 的中分线方向。以此类推, 直至轨迹的倒数第二个点, 简化全部轨迹点。

2) 方位角的特征判断如图 2 所示。其中, T_{OE} 和 T_{OA} 为轨迹 OE 和 OA 的线方位角, 点 A 、 B 、 C 、 D 均满足角度和速度阈值的传递要求, 由于 $|T_{OE} - T_{OA}| > \beta_{\text{max}}$, 表明从 D 点之后所有点超过方位角阈值, 因此, 将超过方位角阈值 E 点的上一个点 D 视为轨迹的方位角特征点, 最终简化轨迹为 ODE , 即 OD 段和 DE 段轨迹; 当 $2\beta_{\text{max}} = 360^\circ$ 时, 不设置方位角阈值, 此时算法为整体轨迹的简化; 当 $2\beta_{\text{max}} < 360^\circ$ 时, 算法存在方位角阈值判定, 为分段轨迹的简化。

1.4 阈值引导采样法算法流程图

算法流程如图 3 所示。首先输入一条轨迹的特征点集, 对数据点进行中断特征点的判定, 将轨迹

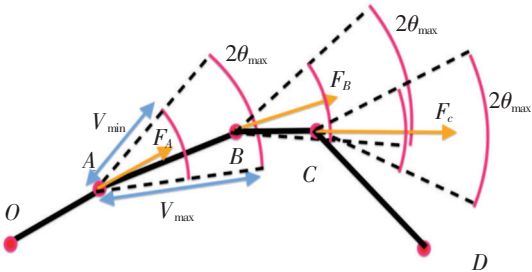


图 1 角度与速度阈值

Fig.1 Angle and speed thresholds

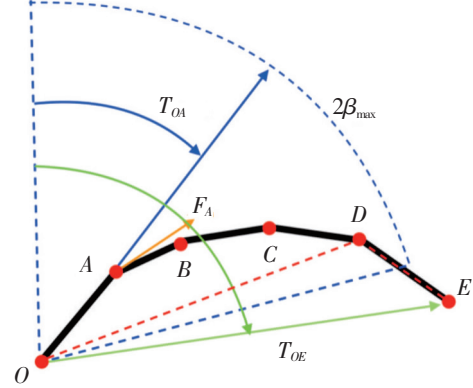


图 2 方位角阈值

Fig.2 Azimuth threshold

进行分割；其次，对数据点进行方位角特征判定，对轨迹进行分段；最后，通过判断角度与速度特征进行子轨迹的简化。

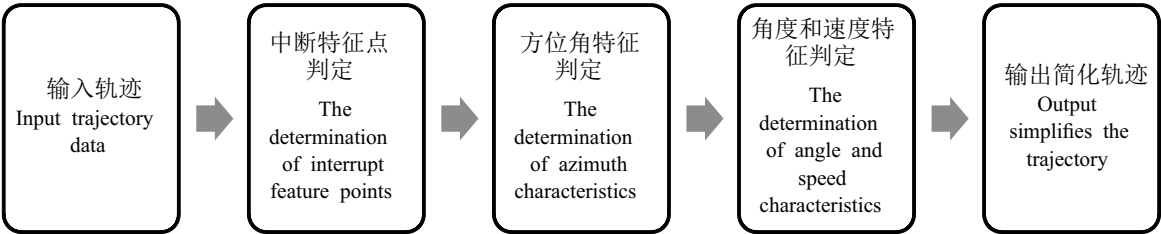


图 3 阈值引导采样算法流程图

Fig.3 Flow chart of threshold guided sampling algorithm

1.5 阈值引导采样法的伪代码

输入原始轨迹 $T = \{T_1, \cdots, T_i, \cdots, T_N\}$, $T_i = \{p_1, \cdots, p_j, \cdots, p_n\}$, $p_j = (i, s_j, c_j, t_j, \text{lon}_j, \text{lat}_j)$; 输入速度阈值 S_{\max} 、角度阈值 θ_{\max} 、方位角阈值 β_{\max} 、时间间隔阈值 t_{\max} 。简化后的轨迹段集合为: $L_{\text{simp}} = \{P_{\text{simp_}1}, \cdots, P_{\text{simp_}m}, \cdots, P_{\text{simp_}k}\}$, $(1 \leq m \leq k)$ 。其中简化的轨迹段 $P_{\text{simp_}m}(p_{m_{\text{start}}}, p_{m_{\text{down}}})$ 元素 $p_{m_{\text{start}}}$ 和 $p_{m_{\text{down}}}$ 分别为轨迹点集简化后选取的特征点的起点和终点。

基于阈值引导采样法的船舶轨迹简化算法如下：

- 1) 输入：原始轨迹 T , 参数 S_{\max} 、 θ_{\max} 、 β_{\max} 、 t_{\max} ,
- 2) for i in N ,
- 3) for j in n :
- 4) $t_{\text{diff}} = |t_j - t_{j-1}|$, $S_{\text{diff}} = |s_{j-1} - s_j|$, $C_{\text{diff}} = |c_{j-1} - c_j|$,
- 5) if $t_{\text{diff}} > t_{\max}$:
- 6) $p_{j-1} = p_{m_{\text{down}}}$, $m = m + 1$, $p_j = p_{m_{\text{start}}}$,
- 7) continue ,
- 8) Else if $|azi_{i+1} - azi_{ij}| > \beta_{\max}$:
- 9) $p_{m_{\text{down}}} = p_{j-1}$, $m = m + 1$, $p_{m_{\text{start}}} = p_{j-1}$,
- 10) Else $C_{\text{diff}} > \theta_{\max}$ or $S_{\text{diff}} > S_{\max}$:
- 11) $p_{m_{\text{down}}} = p_{j-1}$, $m = m + 1$, $p_{m_{\text{start}}} = p_{j-1}$,
- 12) 输出：轨迹段集合 L_{simp} 。

2 实验结果分析

本文选取了两组实验数据与角度阈值法进行对比分析，为了避免误差分析的不准确，轨迹的预处理条件均相同，采用的验证聚类算法为 DBSCAN^[18]，轨迹相似性度量的方法选择为 Hausdorff 距离^[17]。其中，DBSCAN 的参数 Eps（领域距离阈值）和 MinPts（密度阈值）为反复多次实验选取。实验所有涉及距离和方位角信息均以 WGS84 为坐标进行计算。

第一组轨迹简化和聚类实验，选取了厦门港 2012 年 1 月 1 日全天的 AIS 数据，一共有 88 条船舶，3 650 个数据点。水域范围选取 24. 340°N ~ 24. 450°N，118. 130°E ~ 118. 167°E；方位角阈值参数 $\beta_{\max} = 180^\circ$ ；时间间隔阈值 $t_{\max} = 180\text{ s}$ ，180 s 为 AIS 信息播发的最大间隔。同时为了使初始数据具有可操作性与真实性，删除了以下几种情况的 AIS 数据：MMSI ≤ 100000000；0° > COG；COG > 360°；SOG > 50 kn；SOG < 0 kn。

第二组简化分段聚类实验，选取了厦门港 2016 年 1 月 1 日至 2016 年 1 月 13 日的 10 000 条船舶的 AIS 数据点，水域范围选取 24. 350°N ~ 24. 600°N，117. 900°E ~ 118. 150°E，船舶类型选取船长大于 230 m 的货船。各参数值为：时间间隔阈值为 180 s；方位角阈值设置为 15°；因船舶的移动特性，

将角度阈值设置为 10° ; 速度阈值设置为 3 kn; 因厦门港航道宽度约为 300 m, 经反复实验将 DB-SCAN 的 Eps 设置为 230 m, MinPts 为 5。

第三组实验与第二组实验数据水域范围相同, 选取 2016 年 2 月 1 日至 2016 年 2 月 9 日的 20 000 条 AIS 数据点, 船舶类型不做限定, 船长选择大于 20 m。各参数取值为: 时间间隔阈值为 180 s; 方位角阈值为 20° ; 因船舶的移动特性, 将角度阈值设置为 10° ; 速度阈值为 3 kn; 经反复实验将 DB-SCAN 的 Eps 设置为 600 m, MinPts 为 5。

2.1 简化率和误差定义

定义 1: 设轨迹简化率 η 为剩余简化轨迹点的数量 $n - p_{a_b}$ 与原始轨迹点数量 n 之间的比值, 即 $\eta = (n - p_{a_b})/n$ 。其中: p_{a_b} 表示角度阈值为 a , 速度阈值为 b 时的简化轨迹点数量; $n \geq 2$, 且 $n \geq p_{a_b}$ 。

定义 2: 设简化效率 μ 为 A 方法轨迹简化率 η_A 与 B 方法轨迹简化率 η_B 的比值, 即 $\mu = \eta_A/\eta_B$ 。

定义 3: 设简化误差率 δ 为 A 方法简化后剩余轨迹与 B 方法简化后剩余轨迹之差与原始轨迹点数之间的比值, 即 $\delta = (p_A - p_B)/n$ 。

2.2 轨迹简化实验结果分析

用第一组实验数据分析船舶轨迹简化结果。原始轨迹如图 4 所示, 该水域包括两条航道和一个锚泊水域, 其中位于水域最顶部较多曲折轨迹为锚泊水域, 中间与下半部分的轨迹集为两条航道。

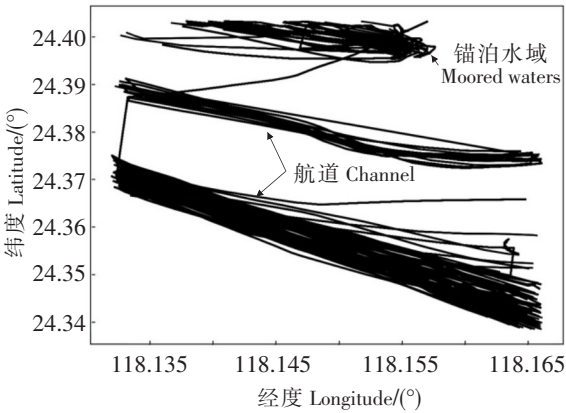


图 4 原始轨迹

Fig.4 Original trajectory

采用控制变量法对各参数的简化效率进行分析, 简化后剩余的轨迹点实验结果如图 5 所示。

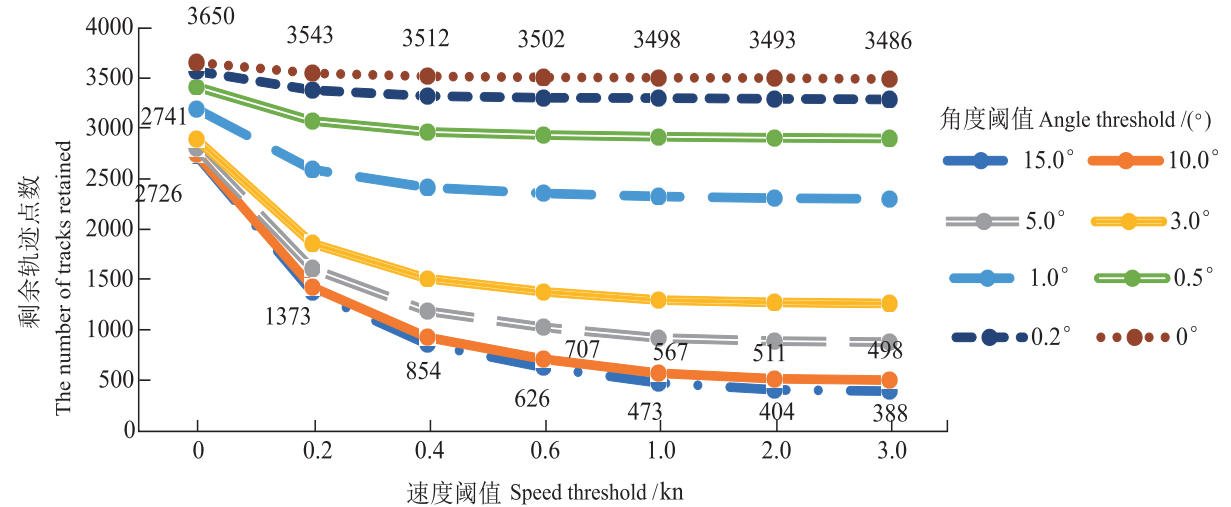


图 5 简化后剩余的轨迹点

Fig.5 Track points remaining after simplification

在进行中断特征点的判定后, $S_{\max} = 0$ 时的阈值引导采样法, 即为 Long^[10] 提出的角度阈值法。采用阈值引导采样法分别对角度阈值的变化以及速度阈值的变化进行实验, 简化结果可知: 当角度阈值固定时, 随着速度阈值的增大, 每一条折线的纵坐标依次递减, 数据减少, 简化效率提高; 当速度阈值固定, 随着角度阈值的增大, 每一列 (相同速度的不同角度阈值视为一列) 的纵坐标依次递减, 数据减少, 简化效率提高。可见本算法比角度阈值法有了较大提升。例如: 实验中选取角度阈值法进行简化, 15° 角度阈值时, 简化率最高为: $\eta = (n - p_{15_0})/n = (3650 - 2726)/3650 = 25.31\%$ 。

以阈值引导采样法进行轨迹简化, 当角度阈值为 15° , 速度阈值为 3 kn, 轨迹简化率 $\eta = (n -$

$p_{15_3})/n = (3650 - 338)/3650 = 90.73\%$ ；简化效率 $\mu = \eta_{15_3}/\eta_{15_0} = 90.71\%/25.31\% \approx 3.6$ 。

即阈值引导采样法的简化效率是角度阈值法的 3.6 倍。同理，当以速度阈值不变，角度阈值变化的情况下，随着角度阈值增大，简化效率也增大，简化剩余的轨迹数越少。

简化效率并非越高越好，而是应该在尽可能保留特征属性的情况下进行高效的简化。在角度阈值为 0° ，速度阈值为 2 kn 与 3 kn 的条件下，简化结果表明，仅存在 7 个数据点的相邻速度变化超过了 2 kn，即速度特征点为 7 个；角度阈值为 15° ，速度阈值为 2 kn 与 3 kn 的条件下，简化结果表明，存在 16 个数据点的相邻速度变化超过了 2 kn，即多识别出 9 个角度特征点。因此相比角度阈值法，阈值引导采样法的简化误差率 $\delta = [(p_{15_2} - p_{15_3}) - (p_{0_2} - p_{0_3})]/n = [(404 - 388) - (3493 - 3486)]/3650 = 2.47\%$ 。

以上分析表明，经过处理后的数据依旧具有较好的特征点识别率，能极大地保留原始轨迹的形状与特征点。

2.3 轨迹简化算法的聚类实验

在轨迹简化的应用方面，阈值引导采样法不仅有较好的聚类效果，而且能精准地识别角度阈值法不能识别的异常轨迹。

第一组实验的轨迹聚类结果如图 6 所示。轨迹集一共分为三个簇（cluster），代表了 3 条习惯航路的轨迹，航迹 A 和航迹 B 被标记为异常轨道。其中：航迹 A 横穿中间航道，该部分原始轨迹局部放大后如图 7 所示；航迹 B 处于中间航道和下侧航道中间，阈值引导采样法的轨迹聚类结果如图 8 所示。除了 3 条习惯航路的轨迹被准确识别外，横穿航道的航迹 A 以及部分航行数据异常的轨迹也被识别出，横穿航道的航迹 A 被进行了准确分割，局面放大后如图 9 所示。

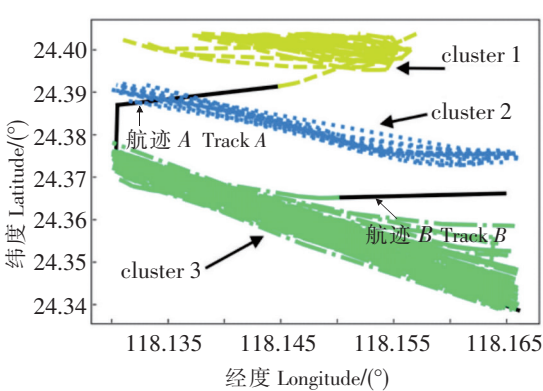


图 6 角度阈值法聚类
Fig.6 Angle threshold method clustering

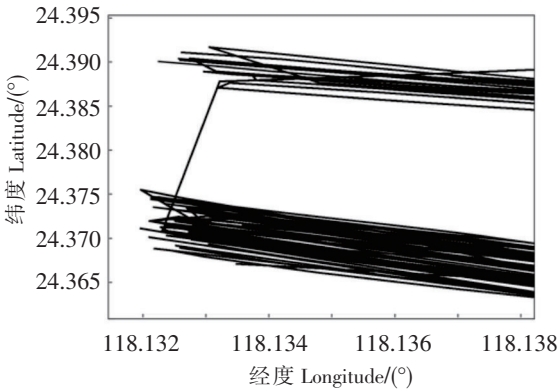


图 7 部分原始轨迹
Fig.7 Part of the original trajectory

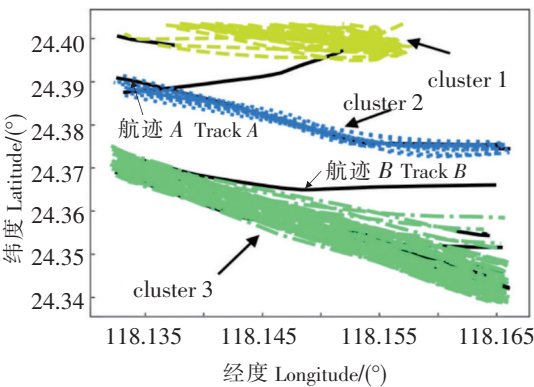


图 8 阈值引导采样法聚类
Fig.8 Threshold guide sampling clustering

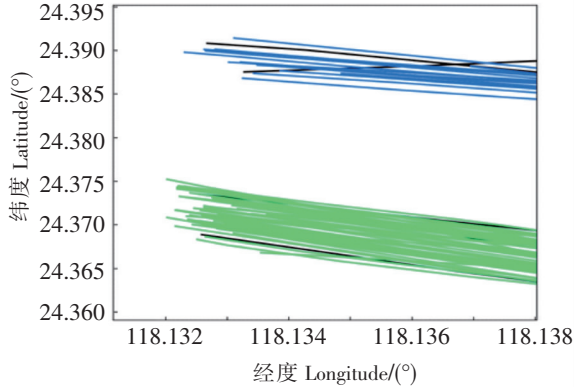


图 9 部分简化后的轨迹
Fig.9 A partially simplified track

综上可知: 对比角度阈值法, 本算法不仅能完全识别横穿航道的轨迹, 而且可以对时间间隔过大的轨迹进行分割, 避免了反复进出研究水域的相同 MMSI 的轨迹点被视为同一时序中的序列, 提高了数据的质量。

第二组实验在区域较大且环境复杂的港口水域进行。该实验不同于整体聚类, 需要将轨迹先进行分段简化再聚类。由于主航道至海沧航道的转向较小, 因此从海沧航道至主航道的来往船舶轨迹视为一类; 而东渡航道至主航道往来船舶在 Y 型分叉口有极大的转向, 因此视为两类船舶轨迹, 一类为东渡航道至主航道往来船舶截至 Y 型分叉口的轨迹, 一类为主航道至东渡航道来往船舶截至 Y 型分叉口的轨迹。

实验的聚类结果如图 10 所示。其中: 簇 1 为从海沧航道至主航道的船舶来往轨迹; 簇 2 即为东渡航道至主航道往来船舶截至 Y 型分叉口的轨迹; 簇 3 为主航道至东渡航道来往船舶截至 Y 型分叉口的轨迹。

实验结果表明: 本文提出的轨迹简化算法对原始轨迹进行简化后, 其简化的轨迹能应用于轨迹聚类, 能分段聚类出较大转向的船舶轨迹。

第三组实验数据为船长 20 m 以上的船舶在厦门港的聚类, 此时船舶轨迹出现在招银航道、厦鼓航道, $t_{\max} = 180\text{ s}$, $S_{\max} = 3\text{ kn}$, $\theta_{\max} = 10^{\circ}$, $\beta_{\max} = 20^{\circ}$, 船舶的轨迹点的数量由 20 000 简化至 3930, 简化率 $\eta = (n - p_{10_3})/n = (20\ 000 - 3930)/20\ 000 = 80.35\%$, 轨迹简化效果较为理想。

实验的聚类结果如图 11 所示。其中: 簇 1 为主航道至招银航道的船舶往来轨迹; 簇 2 为海沧航道至主航道的船舶来往轨迹; 簇 3 为东渡航道的船舶轨迹; 簇 4 为主航道至厦鼓航道的船舶往来轨迹; 簇 5 位为东渡航道至主航道往来船舶截至 Y 型分叉口的轨迹。

实验结果表明: 在频繁变向的船舶轨迹数据中, 本文提出的简化算法简化率良好, 将简化的轨迹数据进行聚类实验, 聚类效果较为理想。

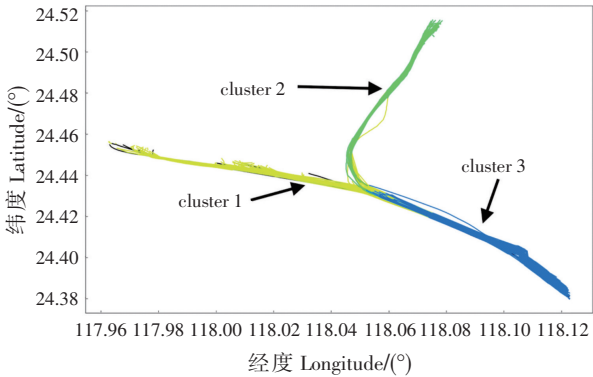


图 10 厦门港船长 230 m 以上的船舶的轨迹聚类
Fig.10 Vessel trajectory clustering of vessels of less than 230 meters in length in Xiamen Port

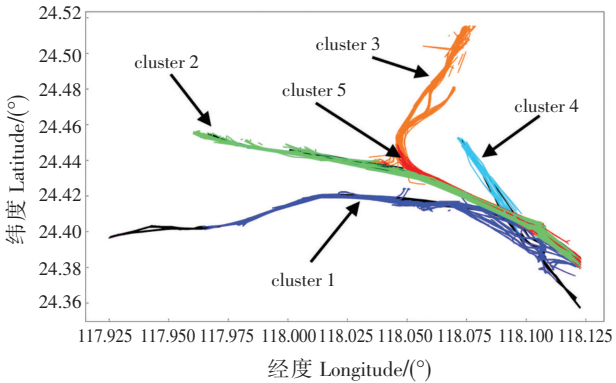


图 11 厦门港 20 m 以上船长的船舶的轨迹聚类
Fig.11 Vessel trajectory clustering of vessels of less than 20 meters in length in Xiamen Port

3 结论

阈值引导采样法包含角度阈值法的特征点判定, 而且综合衡量了轨迹的速度、方向、位置、时间信息。该算法不仅比角度阈值法有更好的简化效率, 并且可以根据不同的阈值选取不同精度的简化结果, 具有更加良好的适应性。

在计算复杂度方面, 阈值引导采样法与角度阈值法一致, 仅需对所有的点遍历一次, 具有高效的计算复杂度。阈值引导采样法简化误差率为 2.47%, 符合实际使用的允许误差要求。在轨迹简化后聚类适应性上, 阈值引导采样算法不仅能进行整体的轨迹简化, 而且还能对轨迹进行简化分段聚类。

[参考文献]

- [1] VESPE M, VISENTINI L, BRYAN K, et al. Unsupervised learning of maritime traffic patterns for anomaly detection [C] //Data Fusion & Target Tracking Conference. London: IET, 2012: 1-5.
- [2] 宋鹏. 基于 C-OPTICS 算法的船舶轨迹聚类与应用 [D]. 大连: 大连海事大学, 2017.
- [3] ZHENG Y. Trajectory data mining: an overview [J]. ACM Transactions on Intelligent Systems and Technology, 2015, 6(3): 1-41.
- [4] JONATHAN MUCKELL, PAUL W Olsen, JEONG-HYON HWANG, et al. Compression of trajectory data: a comprehensive evaluation and new approach [J]. GeoInformatica, 2014, 18(3): 435-460. DOI 10.1007/s10707-013-0184-0.
- [5] LIANGBINZHAO, GUOYOU SHI. A trajectory clustering method based on douglas-peucker compression and density for marine traffic pattern recognition [J]. Ocean Engineering, 2019, 172(12): 456-467.
- [6] BURAGOHAHN C, GANDHI S, HERSHBERGER J, et al. Contour approximation in sensor networks [C] //IEEE International Conference on Distributed Computing in Sensor Systems. Berlin: Springer, 2006, 356-371.
- [7] 潘家财, 邵哲平, 姜青山. 数据挖掘在海上交通特征分析中的应用研究 [J]. 中国航海, 2010, 33(2): 60-62, 73.
- [8] 肖潇. 基于 AIS 信息的船舶轨迹聚类模型研究 [D]. 厦门: 集美大学, 2015.
- [9] 何苗. 移动对象的时空轨迹聚类算法研究 [D]. 兰州: 兰州大学, 2013.
- [10] LONG C, WONG C W, JAGADISH H V. Direction-preserving trajectory simplification [J]. Proceedings of the Vldb Endowment, 2013, 6(10): 949-960.
- [11] 秦育罗, 郭冰, 孙小荣. 改进 Hausdorff 距离及其在多尺度道路网匹配中的应用 [J]. 测绘科学技术学报, 2020, 37(3): 313-318.
- [12] 汪祖云, 张道航, 刘文平, 等. 基于轨迹段核密度的旅游车辆轨迹聚类算法 [J]. 测控技术, 2020, 39(9): 108-112, 117.
- [13] 江玉玲, 熊振南, 唐基宏. 基于轨迹段 DBSCAN 的船舶轨迹聚类算法 [J]. 中国航海, 2019, 42(3): 1-5.
- [14] LEE J, HAN J, WHANG K. Trajectory clustering: a partition-and-group framework [C] //Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, New York: ACM, 2007, 593-604.
- [15] 艾佳琪. AIS 航迹聚类分析及异常轨迹检测 [D]. 大连: 大连海事大学, 2020.
- [16] 胡智辉. 基于轨迹聚类的船舶异常行为识别研究 [D]. 厦门: 集美大学, 2017.
- [17] 赵兴旺, 梁吉业. 一种基于信息熵的混合数据属性加权聚类算法 [J]. 计算机研究与发展, 2016, 53(5): 1018-1028.
- [18] 荣秋生, 颜君彪, 郭国强. 基于 DBSCAN 聚类算法的研究与实现 [J]. 计算机应用, 2004(4): 45-46.

(责任编辑 陈 敏 英文审校 周云龙)