

2021年世界交通运输大会水运学部会议 基于改进 K 中心点聚类的船舶典型轨迹 自适应挖掘算法

李倍莹¹, 张新宇¹, 沈忱², 姚海元², 齐越²

(1. 大连海事大学航海学院, 辽宁 大连 116026; 2. 交通运输部规划研究院, 北京 100028)

摘要:针对目前船舶典型轨迹的挖掘多以轨迹段作为基本单元, 导致聚类对象较为复杂且聚类参数难以确定的问题, 本文提出一种基于改进 K 中心点聚类的船舶典型轨迹自适应挖掘算法。算法以轨迹点作为聚类对象, 分析船舶的航速、航向特征并对轨迹点进行压缩; 将分段均方根误差引入 K 中心点聚类算法, 实现聚类参数的自适应选择; 提取其中的聚类中心点作为轨迹特征点, 得到不同类别船舶的典型轨迹。以天津港主航道船舶自动识别系统 (automatic identification system, AIS) 数据为例, 基于地理信息系统平台 ArcGIS 实现聚类结果的可视化展示。实验结果表明, 运用该算法得到的船舶典型轨迹与实际相符, 自适应程度较高。研究结果对于辅助船舶轨迹异常检测及挖掘海上交通特征具有重要意义。

关键词: 海上交通数据挖掘; 船舶典型轨迹; K 中心点聚类; 轨迹特征点; 自适应

中图分类号: U675.79

文献标志码: A

Meeting of the Waterborne Transport Division, World Transport Convention 2021 (WTC 2021)

Adaptive algorithm for ship typical trajectory mining based on improved K -medoids clustering

LI Beiyang¹, ZHANG Xinyu¹, SHEN Chen², YAO Haiyuan², QI Yue²

(1. Navigation College, Dalian Maritime University, Dalian 116026, Liaoning, China;

2. Transport Planning and Research Institute, Ministry of Transport, Beijing 100028, China)

Abstract: Currently, the trajectory segment is taken as the basic unit in the ship typical trajectory mining, which leads to complex clustering objects and difficulty in determining cluster parameters. To solve the problem, the adaptive algorithm for ship typical trajectory mining based on the improved K -medoids clustering is proposed. The algorithm takes trajectory points as clustering objects. The characteristics of ship speed and course are analyzed, and the trajectory points are compressed; the segmented root mean square error is introduced into the K -medoids clustering algorithm to realize the

收稿日期: 2021-03-25 修回日期: 2021-05-12

基金项目: 国家自然科学基金(51779028)

作者简介: 李倍莹(1997—), 女, 黑龙江鹤岗人, 硕士研究生, 研究方向为海事大数据挖掘, (E-mail)libeiying0413@163.com;

张新宇(1978—), 男, 吉林长春人, 教授, 博导, 博士, 研究方向为海事大数据、船舶交通组织, (E-mail)zhangxy@dlmu.edu.cn

adaptive selection of clustering parameters; the cluster center points are extracted as the trajectory feature points, and the typical trajectories of different types of ships are obtained. Taking automatic identification system (AIS) data of the main channel of Tianjin Port as an example, this paper realizes the visualization of clustering results based on the geographic information system platform ArcGIS. The experimental results show that the ship typical trajectories obtained by the algorithm are consistent with the actual situation, and the degree of self-adaptation is high. The research results are of great significance to assist the detection of abnormal ship trajectories and mine the characteristics of marine traffic flow.

Key words: marine traffic data mining; ship typical trajectory; K -medoids clustering; trajectory feature point; self-adaption

0 引言

随着经济的快速发展,海上交通密度逐年上升,船舶也呈现出大型化、高速化发展趋势,海上通航安全管理变得愈发困难。船舶自动识别系统 (automatic identification system, AIS) 作为一种新型助航系统,能够发送海量的、表征船舶运动状态的数据。对 AIS 数据进行深度挖掘,可发现船舶的典型行为模式,其中,最为直观的船舶行为模式由船舶轨迹体现。对船舶轨迹数据进行聚类分析,挖掘出船舶典型的运动轨迹,可为船舶轨迹异常检测^[1]和航迹预测^[2]提供技术支持,还可用于船舶航行时间预测^[3]、航路规划^[4]和船舶交通流特征分析^[5]等。

目前,船舶轨迹数据聚类分析方法主要有:基于轨迹段的聚类方法和基于轨迹点的聚类方法^[6]。基于轨迹段的聚类方法通过采用相关的聚类算法对船舶轨迹数据进行聚类,得到具有相似运动模式的轨迹簇。根据聚类对象的不同,基于轨迹段的聚类可分为基于轨迹整体的聚类和基于轨迹片段的聚类两种模式。LI 等^[7]将船舶轨迹整体作为聚类对象,用不同轨迹间的距离衡量轨迹之间的相似度;牟军敏等^[8]、曹妍妍等^[9]运用轨迹整体的典型特征对物体的轨迹进行聚类,将改进的 Hausdorff 距离作为度量轨迹整体间相似度的指标,得到最终的聚类结果。LEE 等^[10]首先提出一种基于轨迹段的聚类算法,用最小描述长度原则对轨迹进行分割,然后运用基于密度的线段聚类算法,得到船舶的公共子轨迹;江玉玲等^[11]提出一种改进的轨迹段聚类算法 DBSCAN (density-based spatial clustering of applications with noise),将离散 Fréchet 距离作为轨迹段间相似度的衡量标准,最终得出船舶典型轨迹。以上将船舶轨迹段作为聚类对象的方法虽能够较好地反映轨迹片段的特征,但难以确定轨迹之间的相似度阈值,且对线段进行聚类的算法较为复杂,难以将轨迹片段相连构成船舶典型轨迹。基于轨迹点的聚类方法通常

将 AIS 数据中由经纬度坐标表示的船舶位置散点作为聚类对象,进而划分船舶轨迹的类簇,得到船舶的轨迹特征模式。LIU 等^[12]结合船舶的航速和航向特征,在 DBSCAN 的基础上提出 DBSCANSD 算法,对 AIS 数据中船舶轨迹点进行聚类,提取船舶的主要航行轨迹;王莉莉等^[13]提出基于航迹点特征的时间窗分割算法,并运用 K 均值聚类算法进行聚类得到航空器的中心航迹;王森杰等^[14]以船舶轨迹点为研究对象,提出一种分层建模的船舶轨迹聚类算法,并通过改进道格拉斯-普克算法提取船舶轨迹特征点,实现船舶轨迹点的快速聚类。运用船舶轨迹点进行聚类具有运算效率高、聚类对象简单的特点,但聚类参数的确定也是一大难点。

针对船舶轨迹段聚类算法复杂程度较高的问题,本文将经过压缩后的船舶轨迹点作为聚类对象,简化聚类过程,降低其复杂程度;为解决传统的轨迹点聚类算法自适应程度较低的问题,本文提出一种综合考虑船舶航速、航向特征,结合分段均方根误差,实现自适应地对不同类别船舶的轨迹点进行聚类的典型轨迹挖掘算法。

1 船舶轨迹点数据预处理

经过解码后的船舶 AIS 数据,还需要经过清洗才能使用。清洗过程包括删除重复的 AIS 数据,删除有明显错误的信息,删除不符合常识的异常数据等。为减少计算量,提高算法的运行效率,首先需要构建能够表示船舶轨迹点的数据结构,在此基础上对船舶的轨迹点进行压缩,得到用于船舶典型轨迹挖掘的聚类点集。

1.1 船舶轨迹点数据结构的建立

AIS 数据中蕴含了丰富的表示船舶特性的信息,包括船舶海上移动通信业务识别码 (maritime mobile service identity, MMSI)、船名、船舶呼号、船舶类型、船长、船宽、船舶吃水等静态信息,以及船舶位置 (经度、纬度)、对地船速、对地航向、船首向、接

收时间等动态信息。

船舶航行的轨迹点是由一系列随空间和时间变化的散点表征和确定的,而这些散点可以用上述AIS数据中的部分信息表示。基于此,本文的船舶轨迹点数据结构包括:经度、纬度,表示船舶的位置属性;对地船速,表示船舶的速度属性;对地航向,表示船舶的方向属性;接收时间,表示船舶位置的时间属性;船舶类型、船长、船宽,表示船舶的静态属性。

通过建立船舶轨迹点的数据结构,保留了能够反映船舶轨迹点的信息,剔除了冗余信息,从而提高了轨迹数据的存储效率和船舶典型轨迹的挖掘效率。

1.2 船舶轨迹点压缩

设 $T_{i,j}$ 为在某时刻某船的轨迹信息,其中, i 为该船的MMSI, j 为具有同一MMSI的船舶的轨迹点次序。因此, $T_{i,j}$ 和 $T_{i,j+1}$ 分别表示具有相同MMSI的船舶按照接收时间进行排序后得到的相邻轨迹点信息,其中, t_j 和 t_{j+1} 分别表示相邻轨迹点的发送时间, v_j 和 v_{j+1} 分别表示相邻轨迹点处的航速大小, c_j 和 c_{j+1} 分别表示相邻轨迹点处的航向。船舶相邻轨迹点的航速变化率和航向变化率计算方法如下:

$$\begin{cases} v' = \frac{\Delta v}{\Delta t} = \frac{v_{j+1} - v_j}{t_{j+1} - t_j} \\ c' = \frac{\Delta c}{\Delta t} = \frac{c_{j+1} - c_j}{t_{j+1} - t_j} \end{cases} \quad (1)$$

航速变化率和航向变化率的示意图见图1。

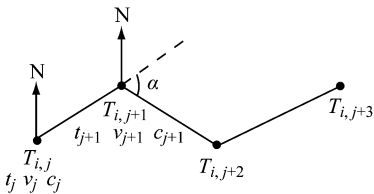


图1 航速变化率和航向变化率示意图

在对船舶轨迹点进行压缩之前,需要确定航速变化率和航向变化率的阈值。若选择的阈值太小,则会导致轨迹点数量过多,在进行数据存储时会浪费一定的空间,且算法的运行时间会增加;若选择的阈值太大,则会导致轨迹点数量过少,用于聚类的点较少,难以保证聚类结果的准确度。因此,一般将航速变化率和航向变化率的阈值设为各自的平均值^[15],分别记为 $\overline{v'}$ 、 $\overline{c'}$ 。

船舶轨迹点压缩的流程如下:

- (1) 分别将每艘船轨迹的起点、终点作为备选点。
- (2) 计算每艘船轨迹点的航速变化率、航向变化率及各自的平均值。
- (3) 从第二个轨迹点开始,首先判断轨迹点处的航速变化率情况。若 $v' > \overline{v'}$, 则将该点作为备选点,否则改为判断航向变化率情况;若 $c' > \overline{c'}$, 则将该点选为备选点,否则该点不是备选点。依次遍历直至倒数第二个轨迹点。
- (4) 将筛选出的船舶轨迹点依次按照MMSI、接收时间进行排序,得到船舶轨迹备选点的集合,作为压缩后的船舶轨迹点集,用于后续船舶典型轨迹的挖掘。

2 船舶典型轨迹挖掘模型

本文的船舶典型轨迹挖掘算法是基于轨迹点的聚类算法,在得到经过清洗和预处理的船舶轨迹点数据后,运用自适应K中心点聚类算法进行聚类,通过将传统的K中心点聚类算法与分段均方根误差评价指标相结合,自适应寻找局部最佳K值,最终将聚类中心点相连得到船舶典型轨迹,具体的技术路线见图2。

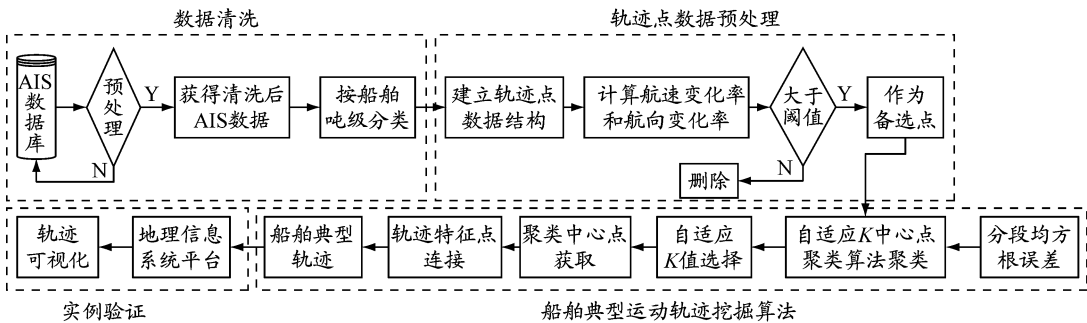


图2 基于轨迹点自适应聚类的船舶典型轨迹挖掘技术路线

2.1 K中心点聚类算法原理

K均值聚类算法和K中心点聚类算法是两种典型的基于划分的聚类方法。当样本数据中存在噪声

点时,使用K均值聚类算法进行聚类会导致结果产生一定的误差。为削弱噪声数据对聚类结果的影响,本文采用K中心点聚类算法对船舶轨迹特征点

进行聚类。该算法步骤与 K 均值聚类算法步骤大体相似,区别在于: K 均值聚类算法通过计算当前类簇中所有数据点的平均值来确定聚类中心,而 K 中心点聚类算法将聚类中心选为当前类簇数据点中的一点。

算法所需的定义如下:

定义 1 任意两样本点 $i(x_i, y_i)$ 与点 $j(x_j, y_j)$ 之间的距离 d 采用欧几里得距离公式进行计算:

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

式中: $i, j = 1, 2, \dots, n$ 。

定义 2 聚类的绝对误差 E :

$$E = \sum_{m=1}^w \sum_{P \in c_n} \overline{PO}_m^2 \quad (3)$$

式中: O_m 为选取的聚类中心点, $m = 1, 2, \dots, w$; P 为类簇 c_n 中的样本点。

具体的算法流程^[16]如下:

步骤 1 从样本数据中随机选取 w 个初始对象 ($O_1, O_2, \dots, O_w$) 作为初始聚类中心点。

步骤 2 根据式(2)计算其他样本点到各聚类中心点的距离,将各样本点划分到与各聚类中心点最近的类簇中。

步骤 3 随机选择一个非中心点 O_r ,使得根据式(3)计算得到的绝对误差减小,不断用 O_r 代替 O_m 。

步骤 4 重复执行步骤 2 和步骤 3,直至计算出的相邻两次绝对误差值保持不变,即聚类中心不再发生变化,否则继续执行步骤 3。

步骤 5 算法终止,输出最终的聚类中心点和类簇。

2.2 基于自适应 K 中心点聚类算法的船舶典型轨迹挖掘

由于 K 中心点聚类算法是一种无监督学习方法, K 值的选择受人的因素影响较大。为实现 K 中心点聚类算法中参数的自适应确定,结合概率论与数理统计理论将分段均方根误差作为评价指标,与 K 中心点聚类算法相结合对传统的 K 中心点聚类算法进行改进,形成自适应 K 中心点聚类算法。均方根误差利用观测值与真实值之间的差值计算得到,可以反映线段对点的拟合程度,其计算式为

$$E_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{oi} - y_{pi})^2} \quad (4)$$

式中: N 为观测样本的数量; y_{oi} 为观测样本的因变量; y_{pi} 为观测样本的预测值,该值越小,表明拟合程度越好。

均方根误差通常将整个样本点数据作为真实值,导致其不能很好地反映局部样本的拟合效果。基于此,本文提出分段均方根误差,即将误差的计算对象缩小为局部样本点,实现误差的局部最小化,并将其与传统 K 中心点聚类算法相结合,作为基于自适应 K 中心点聚类的船舶典型轨迹挖掘算法的参数确定依据。分段均方根误差计算式为

$$E_{\text{SRMS}} = \frac{1}{K+1} \sum_{l=1}^{K+1} E_{\text{RMS},l} \quad (5)$$

式中: K 为最终确定的聚类中心点数, $K+1$ 为样本被聚类中心点所分成的总段数; l 为段的次序。

运用自适应 K 中心点聚类算法挖掘船舶典型轨迹的伪代码如下:

输入:经压缩的轨迹数据集 T ,最大聚类簇 K_{max}

输出:船舶典型轨迹

```

1: for  $K=2$  to  $K_{\text{max}}$  do
2:   get  $O_r$ //随机生成  $K$  个初始聚类中心点
3:   for  $i=1$  to  $n$  do
4:      $d = \text{compute } d(T_i, O_1)$ //根据式(2)计算各轨迹点到初始聚类中心点之间的欧几里得距离,并将各轨迹点划分到与各聚类中心点最近的类簇中
5:     setlabel $i$  = 1
6:     for  $j=2$  to  $K$  do
7:       new $d$  =  $d(T_i, O_j)$ 
8:       if new $d$  <  $d$  then
9:         setlabel $i$  =  $j$ 
10:         $d = \text{new}_d$ 
11:      end if
12:    compute new $E$ //在每一类簇中,根据式(3)计算各轨迹点之间的绝对误差,将具有最小绝对误差的点作为新的聚类中心点
13:    if new $E$  =  $E$  then
14:      break
15:    else
16:       $E = \text{new}_E$ 
17:    end if
18:  end for
19: end for
20: rank  $O_m$ //将最终的聚类中心点按经度进行排序
21:  $L = K+1$ //根据聚类中心点将船舶轨迹点划分到不同的航段(航段数 =  $K+1$ )
22: compute equation//根据相邻两个聚类中心点的经纬度坐标值计算出每一航段的斜截式直线方程
23: compute  $y_{p,i}$ //根据每一航段的斜截式方程计算出对应航段内观测轨迹点的预测值
24: compute  $E_{\text{SRMS}}$ //根据式(5)分别求取在不同  $K$  值条件下的观测轨迹点的分段均方根误差值
25: if ( $E_{\text{SRMS},K-1} > E_{\text{SRMS},K}$ ) and ( $E_{\text{SRMS},K+1} > E_{\text{SRMS},K}$ )//寻找局部最小分段均方根误差值所对应的  $K$  值,作为自适应  $K$  中心点聚类算法所确定的参数
26:   get  $K$ //为防止  $K$  值过大,导致分段均方根误差值无限趋近

```

于0,造成过拟合现象,本文选择具有局部最优特征的K值

```
27: break
28: end if
29: end for
30: save  $O_m$  to list//将最终的聚类中心存储到数据表中,作为船舶典型轨迹的轨迹特征点
```

船舶典型轨迹挖掘示意图见图3。

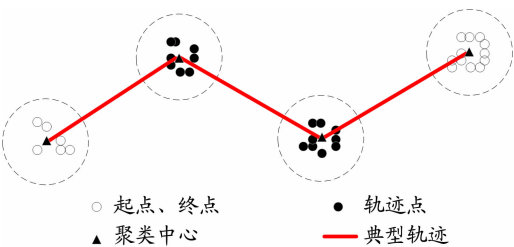


图3 船舶典型轨迹挖掘示意图

3 实验及分析

3.1 AIS数据选取

天津港是我国重要的交通、贸易枢纽,每天进出港船舶较多,交通情况复杂多样,对天津港主航道内的船舶典型轨迹进行研究具有重要意义。选取2019年7月1日至8月31日两个月内进出天津港主航道的货船AIS数据,通过数据清洗筛选出301

艘货船的9 553条AIS数据,以筛选出的数据为研究对象,对船舶典型轨迹挖掘算法进行验证。将船舶吨级按照《海港总体设计规范》(JTS 165—2013)中船长与吨位的对应关系进行划分,将船舶的进出港情况按照航向进行划分^[17],得到不同吨级船舶的AIS数据,见表1。从表1可以看出,经过压缩后轨迹点数量大幅减少(压缩率在58%左右)。

表1 实验所选AIS数据

货船吨级	原始AIS数据/条	压缩后AIS数据/条	进港货船数量/艘次	出港货船数量/艘次
(0.2,1.0]万吨级	1 154	782	64	73
(1.0,3.5]万吨级	1 187	753	58	67
(3.5,10.0]万吨级	3 786	1 975	310	322
(10.0,20.0]万吨级	1 803	1 058	71	69
(20.0,30.0]万吨级	1 623	971	39	33

3.2 聚类参数的选择

将进出天津港主航道的货船按照表1所示的方式对船舶吨级进行划分,运用自适应K中心点聚类算法分别计算K取2,3,⋯,20时的分段均方根误差,从而自适应地确定K值。进港、出港货船分段均方根误差的变化趋势分别见图4和5。

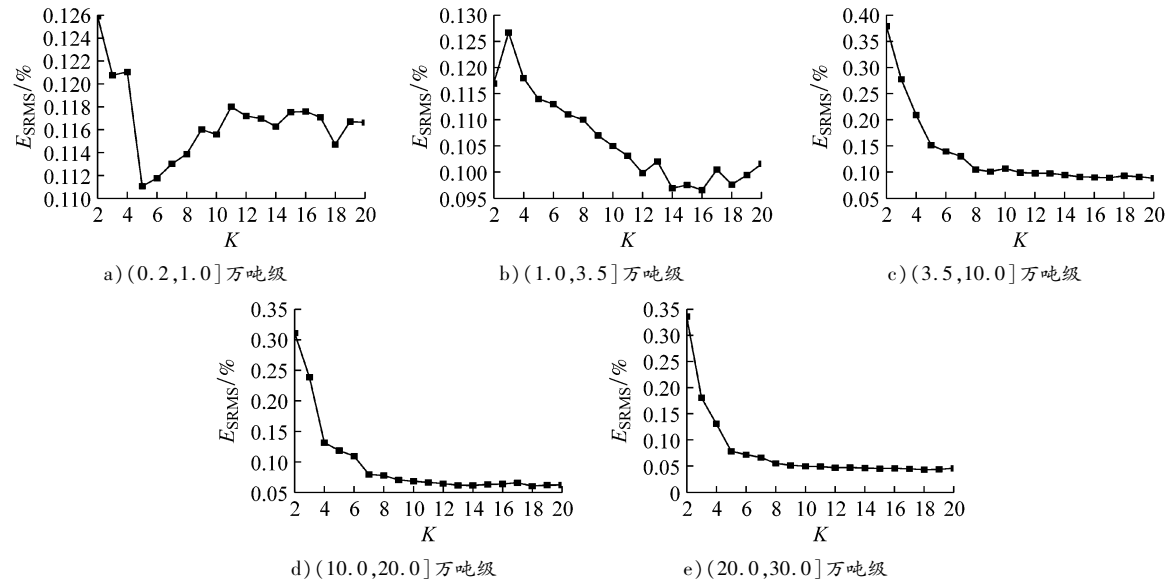


图4 进港货船分段均方根误差变化趋势

从图4和5可以看出,分段均方根误差随K值的增加总体呈下降趋势,这是因为K值越大,用于描述船舶轨迹的特征点越多,总体拟合结果越接近实际船舶轨迹。设分段均方根误差取极小值时的K值为 K_{min} ,当 $K < K_{min}$ 时会出现欠拟合的现象(如图6a所示),当 $K > K_{min}$ 时会出现过拟合的现象(如图6b所示),导致拟合轨迹难以很好地反映真实情况。由于 K_{min} 并不唯一,所以需要确定唯一极小值点,以

相对较少的轨迹特征点较好地反映船舶轨迹特征,从而保留关键特征,减少冗余信息,提高后续研究的计算效率。以(3.5,10.0]万吨级出港货船为例,当K值取11、15和18时,均出现了分段均方根误差的极小值点,因此,在此类船舶的聚类参数进行选择时,将取 $K = 11$ (能使分段均方根误差取极小值的最小K值)作为聚类参数的最优值(如图6c所示)。

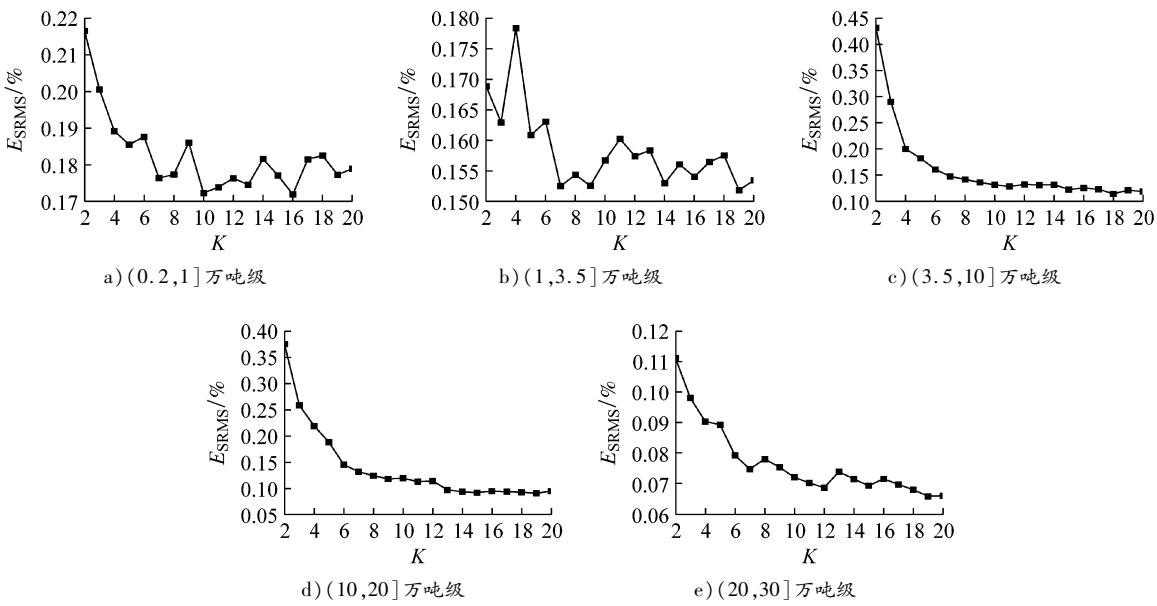


图 5 出港货船分段均方根误差变化趋势

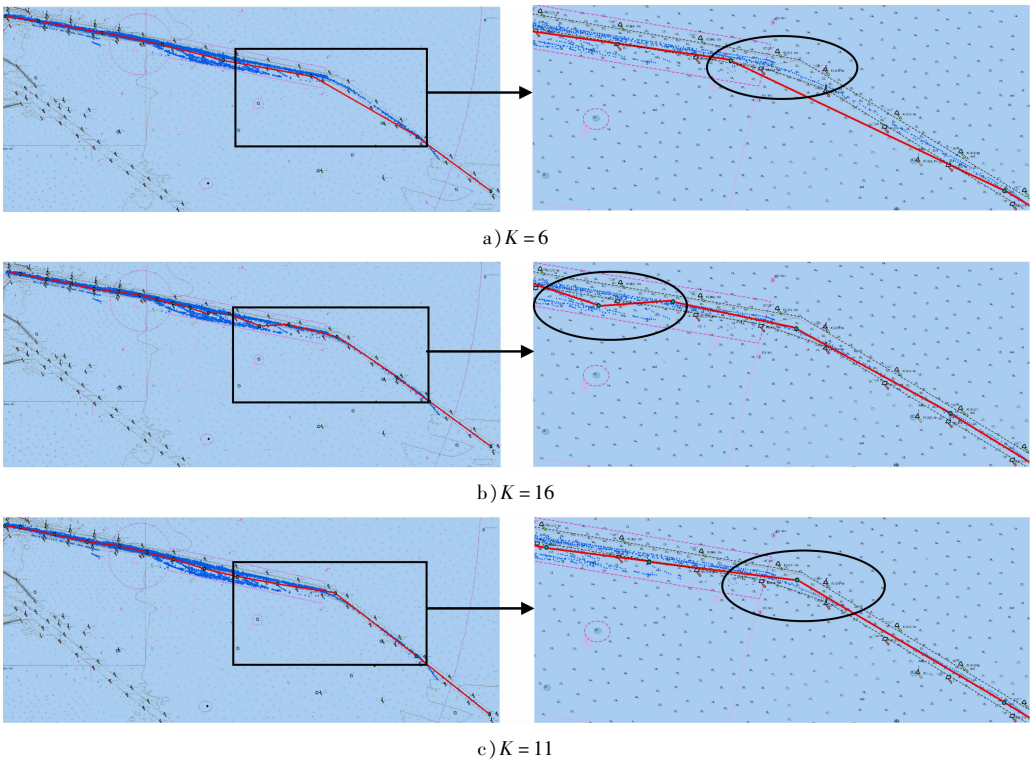


图 6 $(3.5, 10]$ 万吨级出港货船典型轨迹示意图

为验证 $K=11$ 为 $(3.5, 10.0]$ 万吨级出港货船典型轨迹的最佳聚类参数值,用式(3)分别计算在不同 K 值条件下聚类的绝对误差均值,计算结果见表 2。

由表 2 可知,当 $K=11$ 时聚类的绝对误差均值最小,将该值作为聚类参数可以达到最佳的拟合效果。以此类推,在其他分类情况下自适应 K 中心点聚类算法中的 K 值选择情况见表 3。

3.3 船舶典型轨迹挖掘结果

通过分别设置相应的 K 值,运用自适应 K 中心点聚类算法挖掘不同吨级货船进出天津港的典型轨迹,并基于地理信息系统平台 ArcGIS 进行可视化展示(见图 7)。图 7 中,黄色折线表示进港轨迹,红色折线表示出港轨迹。

挖掘结果可以从侧面反映不同吨级的货船在天津港主航道内的通航情况及习惯航路。例如,进港

表2 不同K值条件下聚类的绝对误差均值

K	$E/(\times 10^{-5})$	K	$E/(\times 10^{-5})$	K	$E/(\times 10^{-5})$
2	1.856	9	0.182	16	0.185
3	0.848	10	0.175	17	0.190
4	0.411	11	0.151	18	0.197
5	0.327	12	0.163	19	0.204
6	0.262	13	0.174	20	0.201
7	0.229	14	0.182		
8	0.203	15	0.188		

表3 不同吨级货船轨迹点聚类自适应K值

货船吨级	进港货船	出港货船
(0.2,1.0]万吨级	3	5
(1.0,3.5]万吨级	2	3
(3.5,10.0]万吨级	9	11
(10.0,20.0]万吨级	14	13
(20.0,30.0]万吨级	12	7

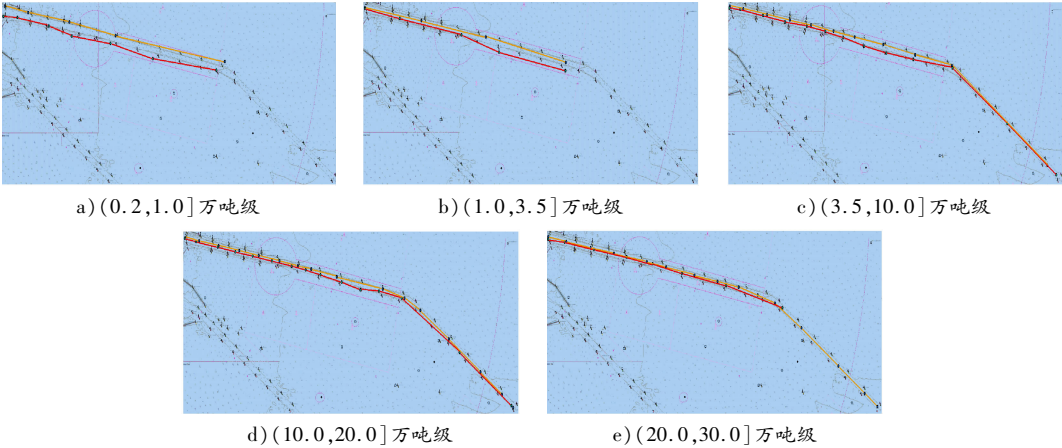


图7 货船进出港天津港典型轨迹

船舶典型轨迹在主航道北侧,出港船舶典型轨迹在主航道南侧,符合靠右行驶的规则。1万吨级及以下的货船在主航道南北两侧的小船航道行驶,3.5万吨级及以下的货船通常从锚地进入航道,而3.5万吨级以上的货船大多从航道入口处进入航道。通过将挖掘出的典型轨迹与实际的天津港航道情况对比可知,二者基本吻合,因此实验结果具有可信度。

为比较本文算法与基于轨迹段的聚类常用的DBSCAN算法的优劣,将运行时间和准确率作为评价指标,基于同一数据集对两种算法分别进行实验。实验环境为: Intel® Core™ i5-3470 @ 3.20 GHz CPU,4 GB 内存,500 GB 硬盘,Windows 10 操作系统,Python 3.6 编程环境。聚类的准确率 R_1 计算式为

$$R_1 = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FP} + N_{FN} + N_{TN}} \tag{6}$$

式中: N_{TP} 表示被正确分类的正例个数; N_{FP} 表示被错分为正例的负例个数; N_{FN} 表示被错分为负例的正例个数; N_{TN} 表示被正确分类的负例个数。 R_1 值越大意味着聚类结果与真实情况越吻合,准确率越高。两种聚类算法的比较结果见表4。

由对比结果可知:自适应K中心点聚类算法在准确率上稍低于DBSCAN算法,其原因为将轨迹段

表4 两种聚类算法的比较结果

算法	运行时间/s	准确率/%
自适应K中心点聚类算法	279	92
DBSCAN 算法	358	96

作为聚类对象使得聚类信息更为丰富,在描述船舶轨迹点特征时具有较好的表现性,但自适应K中心点聚类算法的准确率在90%以上,典型轨迹的挖掘结果可信;在运行时间方面,自适应K中心点聚类算法表现更优,算法的运行效率较高。

4 结 论

本文基于AIS数据,结合船舶的航速、航向特征实现船舶轨迹点的压缩,提高了轨迹点数据的存储效率;将轨迹点作为聚类对象,使得聚类对象简单化,算法运行效率得到提升;根据分段均方根误差值改进了K中心点聚类算法,提高了传统算法的自适应程度;最终选取天津港主航道货船AIS数据,将船舶典型轨迹的挖掘结果基于地理信息系统平台ArcGIS进行可视化展示,结果表明,改进的K中心点聚类算法可有效地对不同类别的船舶轨迹点进行聚类,具有较好的自适应性和准确性。

研究结果可为实现船舶典型轨迹挖掘提供参考,为航线规划、船舶交通行为特征挖掘等提供技术

支持,具有借鉴意义。本文算法在准确率等方面有待进一步提高,运用该结果进一步实现船舶的异常行为检测将是下一步的重点研究方向。

参考文献:

[1] HAN X, ARMENAKIS C, JADIDI M. DBSCAN optimization for improving marine trajectory clustering and anomaly detection[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020, XLIII-B4 – 2020: 455-461. DOI: 10.5194/isprs-archives-XLIII-B4-2020-455-2020.

[2] MURRAY B, PERERA L P. Ship behavior prediction via trajectory extraction-based clustering for maritime situation awareness[J]. Journal of Ocean Engineering and Science, 2021. DOI: 10.1016/j.joes. 2021.03.001.

[3] 唐国磊, 赵宇迪, 于菁菁, 等. 基于 AIS 数据的集装箱船舶到港时间预测研究[J/OL]. 重庆交通大学学报(自然科学版): 1-5[2021-05-09]. <http://kns.cnki.net/kcms/detail/50.1190.U.20181116.0958.008.html>.

[4] 姜大鵬. 基于船舶交通流特征的通航宽度计算及其应用[D]. 大连: 大连海事大学, 2020.

[5] 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.

[6] 张春玮, 马杰, 牛元森, 等. 基于行为特征相似度的船舶轨迹聚类方法[J]. 武汉理工大学学报(交通科学与工程版), 2019, 43(3): 517-521. DOI: 10.3963/j.issn.2095-3844.2019.03.028.

[7] LI Huanhuan, LIU Jingxian, LIU R W, *et al.* A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis [J]. Sensors, 2017, 17: 1792-1817. DOI: 10.3390/s17081792.

[8] 牟军敏, 陈鹏飞, 贺益雄, 等. 船舶 AIS 轨迹快速自适应谱聚类算法[J]. 哈尔滨工程大学学报, 2018, 39(3): 428-432. DOI: 10.11990/jheu.201609033.

[9] 曹妍妍, 崔志明, 吴健, 等. 一种改进 Hausdorff 距离和谱聚类的车辆轨迹模式学习方法[J]. 计算机应用与软件, 2012, 29(5): 38-40. DOI: 10.3969/j.issn.1000-386X.2012.05.011.

[10] LEE J-G, HAN Jiawei, WHANG K-Y. Trajectory clustering: a partition-and-group framework[C]//Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. New York: Association for Computing Machinery, 2007: 593-604. DOI: 10.1145/1247480.1247546.

[11] 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.

[12] LIU Bo, DE SOUZA E N, SYDOW M, *et al.* Knowledge-based clustering of ship trajectories using density-based approach [C]//2014 IEEE International Conference on Big Data. IEEE, 2014: 603-608. DOI: 10.1109/BigData.2014.7004281.

[13] 王莉莉, 彭勃. 航迹点特征的时间窗分割算法的航迹聚类[J]. 空军工程大学学报(自然科学版), 2018, 19(3): 19-23. DOI: 10.3969/j.issn.1009-3516.2018.03.004.

[14] 王森杰, 何正伟. 分层建模的船舶轨迹快速聚类算法[J]. 武汉理工大学学报(交通科学与工程版), 2021, 45(3): 596-601. DOI: 10.3963/j.issn.2095-3844.2021.03.036.

[15] 唐国磊, 赵宇迪, 宋向群, 等. 基于 AIS 数据的集装箱船舶典型运动轨迹研究[J]. 港工技术, 2018, 55(6): 6-9. DOI: 10.16403/j.cnki.ggjs20180602.

[16] ZHEN Rong, JIN Yongxing, HU Qinyou, *et al.* Maritime anomaly detection within coastal waters based on vessel trajectory clustering and Naïve Bayes classifier[J]. Journal of Navigation, 2017, 70(3): 1-23. DOI: 10.1017/S0373463316000850.

[17] 陈向. 复式航道条件下的港口船舶调度优化[D]. 大连: 大连海事大学, 2017.

(编辑 贾裙平)