

AIS 轨迹聚类在港口船舶引航中的研究

石常坡, 朱武斌

(连云港引航站, 江苏 连云港 222042)

摘要: 港口复杂化加重了船舶的管理难度, 为了提高引航安全性, 深度挖掘 AIS 轨迹数据, 对 AIS 数据进行解析和处理, 剔除错误信息, 再利用轨迹聚类模型进行聚类分析, 找出具有相似船舶运动演化方式的轨迹簇, 揭示船舶轨迹内在联系, 本文对传统 DBSCAN 算法进行改进, 加入 AIS 约束条件, 提高聚类精度, 最终实现轨迹的聚类, 从而利用轨迹指导引航工作, 达到安全引航的目的。

关键词: 港口; AIS; 轨迹聚类; 引航

中图分类号: U675.98 **文献标识码:** A

文章编号: 1006—7973 (2022) 10—0141—03

1 引言

近年来, 海洋资源开发和国际航运的高速发展给港口交通和管理提出了更高的要求, 迫切需要提高海上交通信息智能化管理水平, 因此, 基于 AIS 的数据研究成为智能交通研究领域的研究热点, AIS 数据研究在数据挖掘、航行安全、船舶行为分析、贸易分析等领域发挥着重要作用。

AIS 数据作为时空数据类型, 记录着船舶轨迹的位置和时间序列, 通过 AIS 数据挖掘可以识别船舶的航行路线、港口锚泊作业特殊区、港口热点会遇区和碰撞危险区等, 利用 AIS 中蕴含的大量海上交通特征信息, 采用数据挖掘技术, 对船舶轨迹进行聚类分析, 构建轨迹聚类模型, 采用改进的 DBSCAN 算法对大量 AIS 数据聚类研究, 分析船舶进出港口的航行路径、不同季节的交通流方向以及不同类型船舶的引航方法, 为外籍船舶进出港口、主管部门实施航路规划和智能交通研究提供数据支撑。

2 研究水域及数据处理

2.1 研究区域及数据来源

本文研究区域为连云港主港区航道及锚地水域, 研究水域从五号锚地至主港区防波堤, 全长 22 海里, 研究区域内水域开阔无遮挡且附近船舶数量众多, 包括工程船、渔船、商船等, 船舶数量随着国家的发展逐年递增, 船舶引航艘次统计如图 1 所示。根据《国内航行船舶船载电子海图系统和自动识别系统设备规定》: 船载 AIS 设备将不定时的发送动态数据。实验数据选取连云港主航道 2019 年 6 月份的 AIS 数据, 数据包括船舶识别码 (MMSI), 时间 (GPS TIME), 经度 (LONGITUDE), 纬度 (LATITUDE), 速度 (SPEED), 航向角 (COURSE) 等信息, 如表 1 所示, 本文研究船舶类型为 A 级船台, 研究水域年进出船舶量达 6 千多艘次, 为连云港繁忙区。本次研究的 AIS 信息来源于连云港港口采集的 AIS 数据, 从 AIS 数据接收服务器提取出研究水域所需的 AIS 信息

报文文件, 运用 AIS 数据解码模块实现数据解码, 对解码的数据进行数据预处理, 剔除错误信息, 建立 AIS 数据库并对其数据分析, 以使得到的 AIS 数据真实反映船舶航行轨迹。

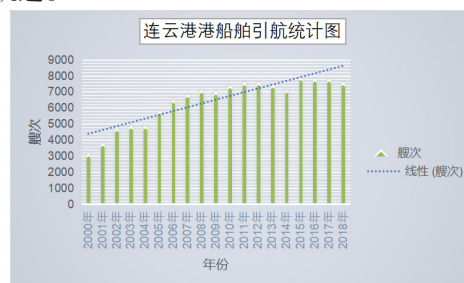


图 1 引航艘次统计

2.2 AIS 数据解析与预处理

船舶 AIS 数据通过船舶 AIS 设备发射到岸上 AIS 基站、地方海事局, 再由相关行业用户进行下载使用, 初始 AIS 数据存在异常数据, 需要对数据进行预处理, 剔除包括航向偏差大、MMSI 位数不对、经纬度不在正常范围内等错误信息, 还包括大量冗余数据, 如果不预处理会存在影响数据准确性、影响运行速度等问题。AIS 数据有 22 种不同类型, 报文分为明暗码两大类, 其中暗码遵照 IEC61162 进行解码, 先对报文进行解析。

然后将解码数据进行预处理, 挑选出所需水域 AIS 信息, 包括动、静态信息, 建立数据库, 最后将两者相同 MMSI 的动静态信息保存到 txt 中, 删除重复数据, 最终提取出船舶的 UTC、MMSI、经纬度、速度、航向等船舶信息。以连云港港口 AIS 数据为例, 数据处理显示为图 2 所示, 其中如图 a 红色航迹为异常数据显示, 图 b 为经过预处理显示。

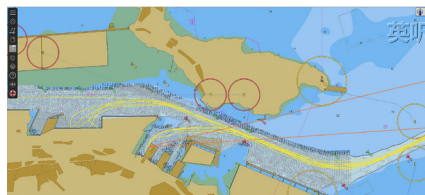


图 a

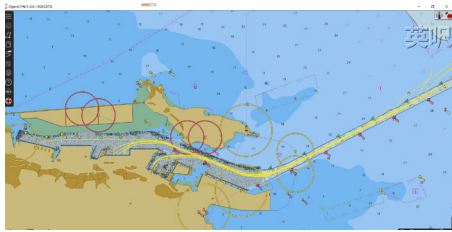


图 b

图 2 AIS 数据轨迹显示

3 AIS 轨迹聚类模型

3.1 主要轨迹聚类算法

船舶轨迹聚类就是利用聚类算法对船舶轨迹进行聚类,找出具有相似船舶运动演化方式的轨迹簇,揭示船舶轨迹内在联系,目前广泛应用于船舶研究的算法大致分为基于距离、密度、统计学三种算法,对应上述算法的典型代表包括: Hausdorff 距离 (Hausdorff Distance, HD) 算法、DBSCAN 算法、混合高斯模型 (Gaussian Mixture Model, GMM) 的轨迹建模方法、K-Means 算法等。

3.2 轨迹聚类总体流程

目前,轨迹聚类方法选取主要从以下两方面进行:

(1) 将整条轨迹作为目标的聚类方法,该方法缺点是轨迹数据量大,时间和空间要求高、轨迹子段丢失等问题;

(2) 将轨迹进行分段划分,对轨迹子段作为目标的聚类方法,运用该方法可能无法完全获取整条轨迹的特征,但能较好地把握轨迹子段的特征,且综合各子段的特征也能较好地对整个轨迹的特征进行描述。

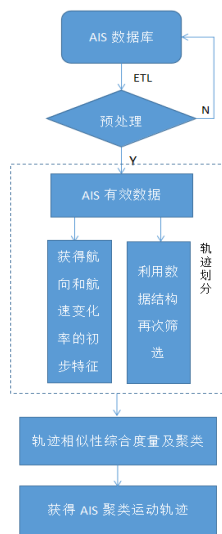


图 3 AIS 轨迹聚类流程图

3.3 轨迹划分

船舶 AIS 信息播发时间间隔根据船舶航向和航速变

化率进行,采集到的 AIS 信息包含大量不同时间信息的轨迹分段,轨迹划分不仅需要体现轨迹的特性,又要保证轨迹的准确性和简洁性,由于连云港航段路程简单,弯曲路段少并且弯曲幅度小,因此船舶 AIS 体现的特征点简单,通过船舶转向的航向角阈值与速度变化率阈值就可以进行划分如图 4。

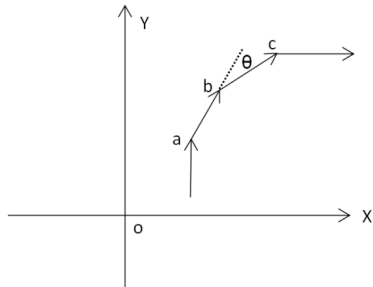


图 4 轨迹划分方案

建立特征点集 (a、b、c....), 将相邻航段之间的航向和航速进行计算,航迹间的夹角为 θ_i , 计算公式如下:

$$\begin{cases} \alpha_v = \frac{|V_b - V_a|}{\Delta t} \\ \alpha_c = \frac{|C_b - C_a|}{\Delta t} \end{cases} \quad (1)$$

式中 α_v 、 α_c 分别为速度变化率和航向变化率, Δt 为相邻时间间隔。

根据上式公式再结合航向航速阈值对轨迹进行分类,将符合阈值的特征点进行收集,阈值设定的设定要兼顾计算速率、细节以及聚类效果。在复杂航段航向航速变化较大的水域,简单的航向阈值和速度阈值不能很好地体现轨迹的原始特征,需要进一步优化特征点的选取,利用智能算法找点全局最优结果,使得数据还原真实轨迹。

3.4 轨迹相似度的度量

船舶轨迹划分需要利用 AIS 数据特征信息进行聚类,如船舶的 MMSI、航向、航速以及船位等 4 个主要特征信息,提高轨迹的聚类效果和准确的,通过对航向、航速、船位进行距离计算,并进行归一化处理,本文还将 MMSI 作为其他三项度量综合权重的系数, MMSI 不符合的轨迹分段直接忽略,减少聚类计算复杂度。相似度的度量主要包括轨迹间距离度量、航速信息度量、MMSI 综合权重系数以及最后的综合度量。根据速度、航向、轨迹距离以及 MMSI 综合权重系数公式,再对相关距离进行归一化处理,使得不同量纲之间可以进行计算,首先定义相关权重 $W = \{W_{//}, W_{\perp}, W_{\theta}, W_s\}$, 且满足权重取值大于等于 0, $W_{//} + W_{\perp} + W_{\theta} + W_s = 1$ 。在对距离进行归一化处理时采用 Z-score 标准法进行,处理过的数据符合正态分布,公式为:

针对狭水道航行船舶的引航研究

张良

(上海港引航站, 上海 200082)

摘要: 狭水道是船舶出入港口、海峡或穿越陆地的重要途径, 但由于狭水道内部建设的复杂性、环境多样以及航线船舶密度大等特点, 导致航行过程中面临多种风险, 因此, 为确保船舶航行安全, 需要专业的引航人员提供相关服务是必要的。随着近年来的船舶大型化、气候多变化、航道通航情况复杂化以及港口发展国际化, 对引航工作的要求逐渐提高。因此更加需要明确狭水道的内部构造以及相关规定, 利用敏锐的观察力及时发现潜在风险和问题。本文针对当前影响船舶在狭水道安全通行的因素进行分析, 探究具有实质性的解决措施。

关键词: 狭水道; 引航; 船舶安全

中图分类号: U675.98

文献标识码: A

文章编号: 1006—7973 (2022) 10—0143—03

航海领域对狭水道的定义较为宽泛, 由于水深、航宽的限制使得船舶在航行过程中受到多方面因素的影响, 环境复杂、通航密度大是制约船舶安全的重点因素, 也是狭水道事故多发的主要原因, 因此, 引航员对船舶的管理和对航道的疏通工作显得十分重要。

1 狭水道概念及特点

迄今为止航海领域对狭水道并没有明确的概念, 一般将狭水道定义为船舶必须靠右行驶才能确保安全的水

域, 但不乏一些国家将狭水道的定义进行数据化处理, 认为航道宽度在 2 海里之内则属于狭水道范围, 随着船舶行业规模的逐渐扩大, 被列为狭水道的水域也逐渐增多。

狭水道顾名思义就是航道比较狭窄、弯路短多且航道内部复杂的水域, 由于狭水道主要位于港口或海峡之间, 因此在这段航线中船舶密度较大, 经常出现拥挤现象, 还存在桥梁、过江缆线、过江隧道、助航器、锚线等建筑或设备, 对船舶的航行带来一定威胁。通常狭水

$$d' = \frac{d - \mu}{\sigma} \quad (2)$$

式 (2) 中 d 为处理前距离, μ 为各距离均值, σ 为标准差。因此利用归一化的距离和权重值求取轨迹聚类的综合距离:

$$D(L_i, L_j) = k(W_d d'_i + W_\perp d'_\perp + W_\theta d'_\theta + W_s d'_s) \quad (3)$$

3.5 建立 AIS 轨迹聚类模型

为进一步加强船舶轨迹聚类效果, 本文对传统 DBSCAN 算法进行改进, 加入 AIS 约束条件, 提高聚类精度, 对基于 AIS 改进的 DBSCAN 中的领域 $\mathcal{E}$, 密度阈值, 船舶航向, 航速以及 MMSI 进行定义。定义轨迹数据集 N 由不超过领域 $\mathcal{E}$ 的轨迹组成, 且定义核心对象满足领域内的轨迹数目大于或者等于密度阈值 $\min S_{tr}$, 定义轨迹 L_i 的数据集 $N (L_i \in N)$, 轨迹点集为 D , p 为轨迹点且 $p \in D$, 轨迹聚类满足航向、航速以及 MMSI 三个条件, 具体定义如下:

$$D_p = \{p \in D | D_{dist}(p_i, p_{i+1}) < \varepsilon\} \quad (4)$$

$$|S_{p_i} - S_{p_{i+1}}| < MaxSog \quad (5)$$

$$|C_{p_i} - C_{p_{i+1}}| < MaxCog \quad (6)$$

$$V_{mmsi} \in \text{相同编号} \quad (7)$$

其中短距离范围为地球曲度可以忽略不计, 在航

向计算过程中需要注意临界数值的转换。

4 总结

船舶 AIS 信息为船舶航行安全提供了数据支持, 通过对 AIS 数据的处理, 利用 AIS 轨迹聚类模型对有效数据进行聚类, 形成有效的航路轨迹, 对不同船舶的航行习惯、路径进行分析, 引航员通过大量的 AIS 数据分析能直观有效地指导引航员操纵船舶, 避开浅滩或者流急水域, 船舶航行的转向时机、转向速度、转头速率等进行学习, 引航员能够直观感受船舶的运动态势, 达到提高引航员的引航安全性。

参考文献:

- [1] CHEN J, LU F, PENG G. A quantitative approach for delineating principal fairways of ship passages through a strait[J]. Ocean eng. 2015, 103: 188-197.
- [2] 李东枫. 基于 AIS 大数据的船舶危险会遇热点区域挖掘研究[D]. 华南理工大学, 2017.
- [3] 徐良坤, 任律珍, 周世波. 船舶 AIS 轨迹聚类方法研究进展综述[J]. 广州航海学院学报, 2019, 27(02): 7-12+47.
- [4] 肖潇, 邵哲平, 潘家财, 纪贤标. 基于 AIS 信息的船舶轨迹聚类模型及应用[J]. 中国航海, 2015, 38(02): 82-86.