

基于 AIS 数据的船舶会遇挖掘与分析

赵吟秋, 索永峰, 鲜 波

(1. 集美大学航海学院, 福建 厦门 361021; 2. 集美大学航海技术研究所, 福建 厦门 361021)

[摘要] 为实现在 AIS 海量数据中快速高效地识别会遇船舶信息, 对会遇船舶航行过程实施监控和分析, 识别研究水域内存在的海上交通安全风险点。以台湾海峡部分区域内的 AIS 数据作为研究对象, 运用改进的 DBSCAN 聚类算法, 结合船舶会遇特征参数计算, 挖掘两船会遇与多船会遇信息, 根据船舶方位和航向差划分船舶交叉、对遇和追越三种会遇局面, 并通过定量化方式多维度描述不同类型的船舶会遇过程。结果表明: 分析轨迹数据的时空特性可以再现船舶会遇场景, 能更清晰全面地展现船舶的会遇过程。

[关键词] 船舶自动识别系统; DBSCAN 聚类; 船舶会遇; 时空特性分析

[中图分类号] U 676.1

Mining and Analysis of Ship Encounters Based on AIS Big Data

ZHAO Yinqiu, SUO Yongfeng, XIAN Bo

(1. Navigation College, Jimei University, Xiamen 361012, China;

2. The Navigation Technology Institute, Jimei University, Xiamen 361012, China)

Abstract: In order to realize the rapid and efficient identification of encountering ships in the massive data of AIS, the research monitors the navigation process of encountering ships by using an improved DBSCAN clustering algorithm, which is well used in rapid extraction of maritime traffic safety risks in the research water area. Through the combination of ship encounters feature parameter calculations and using AIS data within the Taiwan Strait area as the research object, dual ship encounters and multi-ship encounters are mined effectively. Then the encounter situations are divided into three types, ie, cross encounters, head-on encounters and overtaking encounters, based on the ships' bearing and course differences. Finally, the encounter processes of different types of ships are described in multiple dimensions, the encounter scenes of the ship are reproduced through a quantitative method and time-space characteristics analysis. The results show the analysis of time-space characteristics can reproduce ships encounter situations and illustrate the encounter process more clearly and comprehensively.

Keywords: AIS; DBSCAN clustering; ship encounters; analysis of temporal and spatial characteristics

0 引言

船舶自动识别系统 (automatic identification system, AIS) 提供了丰富的船舶航行数据, 基于 AIS 船舶轨迹数据进行通航水域风险评价、船舶自身行为特征分析、自动避碰以及船舶会遇等方面的研究

[收稿日期] 2021-05-16

[基金项目] 国家自然科学基金项目 (51579114)

[作者简介] 赵吟秋 (1998—), 女, 硕士生, 从事交通运输工程信息及控制研究。通信作者: 索永峰 (1981—), 副教授, 博士, 主要从事载运工具运用工程研究。E-mail: yfsuo@qq.com

<http://xuebaobangong.jmu.edu.cn/zkb>

是当前的热点^[1]。潘家财等^[2]根据厦门湾内水域特点以两船会遇距离作为会遇判定条件, 统计船舶会遇的空间分布以及会遇船长分布等, 分析了厦门湾船舶会遇状况。任亚磊等^[3]基于构建的动态船舶领域模型, 对会遇类型、会遇船舶特征等进行统计, 进而评估相关水域的宏观碰撞危险度, 该方法对开阔水域船舶目标航行数据较少的情况下难以确定每条船舶的动态领域, 因此在开阔水域中该方法的适用性还有待验证。马杰等^[4]采用时空关系对数据进行约束, 提取特定时间段内船舶的位置和航向特征, 形成会遇特征序列, 利用支持向量机 (SVM) 对会遇特征进行分类辨识, 但该方法并未列举判定会遇态势的筛选条件, 也不适用于对大量数据进行船舶的会遇挖掘。甄荣等^[5]提出运用墨卡托算法计算船舶之间的距离, 并通过 DBSCAN 算法识别船舶会遇。江龙晖等^[6]根据船舶 AIS 动态数据, 结合插值方法计算船舶间相对运动参数, 提取船舶会遇信息。Du 等^[7]通过检测船舶会遇, 进而对船舶可操纵余量, 避让有效性等指标对会遇船舶的碰撞风险等级进行划分, 检测近距离会遇的船舶。Zhang 等^[8]通过对会遇的船舶进行风险排序, 考虑船舶领域和船舶尺寸计算碰撞风险, 检测近距离会遇的船舶, 该方法可以有效区分具有不同碰撞风险的会遇船舶, 但缺少对会遇过程的量化分析。Kim 等^[9]运用分布式随机搜索算法, 结合 AIS 数据实现会遇的多船各自判断会遇态势。Chen 等^[10]运用改进的时间离散非线性速度障碍 (TD - NLVO) 算法识别特定区域 AIS 数据中的多船会遇局面。

综上所述, 根据海量 AIS 数据快速高效挖掘船舶会遇信息, 定量多维度分析船舶的会遇过程方面的研究有待提高。为了能够快速有效地挖掘船舶会遇局面, 识别潜在的碰撞风险点, 监测处于会遇过程中的船舶航行数据, 定量分析船舶会遇过程, 本文改善传统 DBSCAN 算法, 分析船舶会遇时空演化特性, 有效识别会遇船舶, 并在此基础上挖掘多船会遇局面, 划分船舶不同会遇类型, 分析船舶的会遇避让方式, 建立会遇船舶数据库。

1 AIS 数据预处理

在现实条件下, 由于各种原因导致 AIS 数据存在缺失、重复等数据异常现象, 给 AIS 数据处理以及分析带来一定的难度。为提高数据的准确性和可信度, 首先对 AIS 数据中船舶位置、航速、航向等不符合常理的异常数据进行删除。由于在船速较低的情况下船舶的操纵性能较差, 难以实施有效的避让行为, 分析会遇过程没有实际意义, 因此, 删除船速值小于 4 kn 的 AIS 数据。当 AIS 数据中数据条目的时间间隔较大时, 容易导致后续插值补全的误差很大, 影响再现船舶真实航行情景, 因此, 删除数据条目时间间隔大于 30 min 的数据。

为了对海量的 AIS 数据进行有效分析, 将 AIS 数据条目大于 300 条的船舶航行数据提取出来, 对船舶数据进行后续统计分析, 处理后的部分 AIS 数据如表 1 所示。

表 1 AIS 数据处理项
Tab. 1 AIS data processing items

海上移动通信业务 标识号 MMSI	接收时间 Postime/s	航向 Course/(°)	速度 Speed/kn	经度 Latitude/(°)	纬度 Longitude/(°)	船首向 Ship heading/(°)
813 021 827	1 528 143 873	86.8	6.8	117.451 4	23.601 22	88
813 021 827	1 528 143 875	83.8	5.8	117.451 4	23.601 22	85
201 801 104	1 580 982 016	156.6	7.8	119.085 2	23.817 32	150
568 767 867	1 528 782 369	335.2	19.5	117.559 6	23.674 59	334
813 021 827	1 528 143 933	88.7	6.6	117.453 4	23.601 25	89

2 会遇挖掘算法模型与特征参数计算

2.1 船舶潜在会遇的定义

人们将海上航行船舶相遇而必须采取行动的情况称为“会遇”。会遇不包括无需采取避让行为的

船舶相遇的情况。在海上交通研究中,会遇是指与碰撞事故密切相关且在一定程度上表征海上交通危险的一种特殊的船舶相遇情况^[8]。本文将潜在会遇定义为:当两船位于某一给定距离以内的一种实际情况。参考文献[12]有关船舶最小安全会遇距离的模糊确定研究,本文采用 0.5 n mile 为潜在会遇判定的阈值条件。

2.2 DBSCAN 算法原理

DBSCAN 聚类算法应用广泛^[1],其基本原理如图 1 所示。在研究区域内,设定两个参数阈值 MinPoints 和领域半径 E ,以数据对象 S 为圆心,划定一个半径为 E 的圆形区域,在此区域内检测数据样本数量 Q ,若 $Q \geq \text{MinPoints}$ 则 S 为核心对象,其余点称为边界对象,处于核心对象 E 领域中的样本点为核心对象的直接密度可达样本点;如若 $Q < \text{MinPoints}$ 则数据对象为噪声点,无法隶属任何簇。

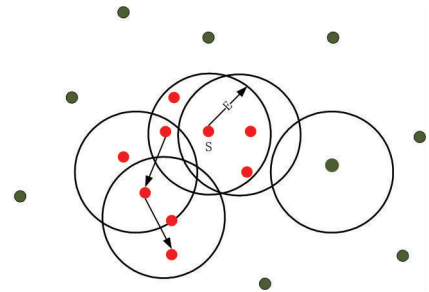


图 1 DBSCAN 聚类

Fig.1 DBSCAN clustering

2.3 船舶会遇特征参数计算

以时间戳为关键字提取船舶航行数据,建立时空约束下不同船舶的轨迹状态时间序列 $\text{traj} = \{(p_1, t_1), (p_2, t_2), (p_3, t_3), \dots, (p_n, t_n)\}$, $p = \{\text{mmsi}, \text{postime}, \text{lon}, \text{lat}, \text{speed}, \text{course}, \text{heading}\}$ 表示标志码为 mmsi 的船舶在某一时刻的航行状态数据,为了避免应用 DBSCAN 算法挖掘船舶会遇时,出现船舶 A 与船舶 B 之间的相对距离值 $d_{AB}(A, B = 1, 2, \dots, n)$ 重复计算的情况,构建船舶相对距离矩阵 D ,计算不同船舶之间的距离^[13]。该矩阵与全矩阵计算相比较,将算法的运算量减小了 50%。

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ 0 & 0 & d_{23} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_{(n-1)n} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (1)$$

给定两船经纬度坐标 $A(\text{Lon}_1, \text{Lat}_1)$, $B(\text{Lon}_2, \text{Lat}_2)$, 采用恒向线方位距离计算公式计算两船方位角 C 和距离 d ^[11]。具体计算公式如下:

$$C = \arctan(\Delta \text{Lon} / \Delta q) \quad (2)$$

$$d = \begin{cases} (S_{\text{Lat}_2} - S_{\text{Lat}_1})^2 \sqrt{\Delta q^2 + \Delta \text{Lon}^2} \sec C / \Delta q, & \Delta \text{Lat} \neq 0, \Delta \text{Lon} \neq 0; \\ r_1 |\Delta \text{Lon}|, & \Delta \text{Lat} = 0; \\ |S_{\text{Lat}_2} - S_{\text{Lat}_1}|, & \Delta \text{Lon} = 0. \end{cases} \quad (3)$$

式中:经度差 $\Delta \text{Lon} = \text{Lon}_2 - \text{Lon}_1$; Δq 为 AB 间纬度渐长率差值; S_{Lat_i} 为点到赤道的子午线弧长; r_1 为 Lat_1 纬度处的等纬圈半径。其中, S_{Lat_i} 的计算式为:

$$S_{\text{Lat}_i} = a(1 - e^2) \int_0^{\text{Lat}_i} (1 - e^2 \sin^2 \text{Lat}_i)^{-3/2} d\text{Lat}_i, \quad i = 1, 2. \quad (4)$$

式中: a 为地球椭圆柱体长半轴长; e 为偏心率。

A 、 B 间纬度渐长率差值 Δq 计算公式如下:

$$\begin{cases} \Delta q = q_2 - q_1, \\ q_i = \ln[\tan(\pi/4 + \text{Lat}_i/2)((1 - e \sin \text{Lat}_i)/(1 + e \sin \text{Lat}_i))^{e/2}] \end{cases} \quad (5)$$

Lat_1 纬度处的等纬圈半径 r_1 为:

$$r_1 = a \cos \text{Lat}_1 / \sqrt{1 - e^2 \sin^2 \text{Lat}_1} \quad (6)$$

本文在两船会遇起始点计算两船航向差与相对方位关系,并将此作为依据划分船舶的会遇类型,

两船航向差 ΔH 以及目标船的相对方位计算如下。

$$\begin{cases} A_0 = C - H_0 \\ \Delta H = |H_0 - H_T| \end{cases} \quad (7)$$

式中: A_0 为目标船的相对方位; H_0, H_T 分别为本船和目标船船首向; ΔH 为两船的航向交叉的角度。

2.4 改进的船舶会遇挖掘算法流程设计

首先对原始 AIS 数据进行预处理, 按照位置时间作为关键字进行升序排序, 以 1 min 为时间间隔提取时间段内船舶航行数据, 将该数据按照船舶识别号分别进行插值补全到同一时刻 T_1, T_2, \dots, T_n , 且每分钟内每条船只保留一条 AIS 数据。遍历每一个时间片, 计算相对距离矩阵, 调用 DBSCAN 算法。DBSCAN 聚类算法设计的关键参数为 $\text{MinPoints} \geq 2$, $E = 0.5$, 以相对距离矩阵 D 为算法输入数据, 输出潜在会遇船舶状态信息, 并将其存储在对应的数据库中, 运用 CPA 算法, 验证数据库中的船舶会遇情况, 最后结合船舶会遇分类划分船舶会遇类型, 进行会遇过程分析。算法流程如 2 图所示。

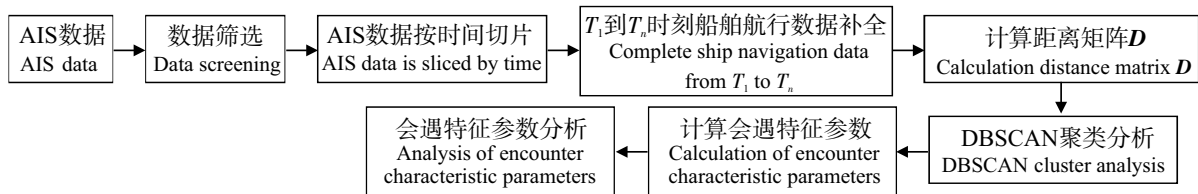


图 2 船舶会遇挖掘算法流程图

Fig.2 Flow chart of ship encounter mining algorithm

3 算法实验

3.1 DBSCAN 聚类实验结果

以 2020 年 6 月台湾海峡的 AIS 数据为研究对象, 原始 AIS 数据为 1000 万条, 经过数据预处理后剩余 400 万条。通过设置滑动时间窗口以提取研究水域内会遇场景的特征序列数据。根据算法流程和参数计算公式, 可以得出某一时刻 T 的潜在会遇船舶信息, 如图 3 所示。其中: 红色点 (-1) 表示没有达到会遇筛选条件的船舶, 即为噪声点; 图 3 中其余颜色 (标识分别为 1, 2, ..., 8) 各自代表一组潜在会遇船舶, 通过标记可以得到潜在会遇船舶的标识码和相关航行状态信息。

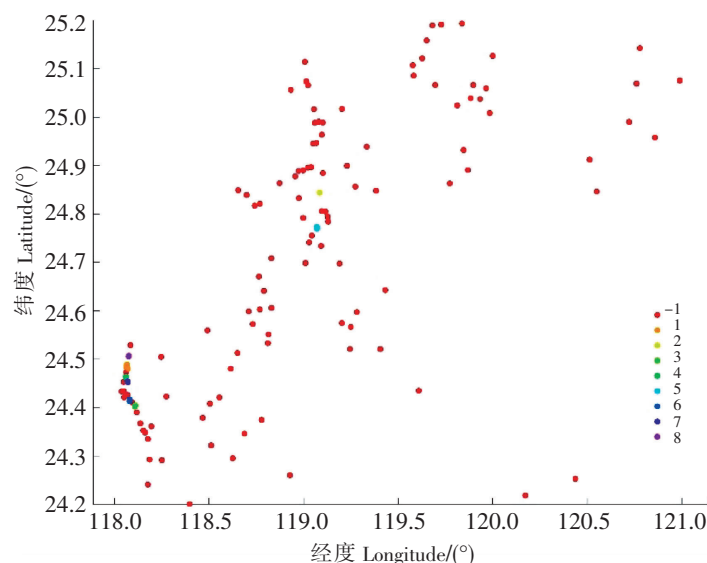


图 3 DBSCAN 算法计算结果

Fig.3 DBSCAN algorithm calculation result

将图 3 部分类别数据导出, 并去除噪声点, 得到表 2。将表 2 中数据进行相关参数计算, 可得两船的相对距离和方位以及航向差信息。

表 2 潜在会遇船舶状态数据
Tab.2 Encounter ship status data

类别 Class	海上移动通信业务 标识号 MMSI	经度 Latitude/(°)	纬度 Longitude/(°)	速度 Speed/kn	航向 Course/(°)	船首向 Ship heading/(°)
class 1	413 696 010	118.062 0	24.483 9	5.5	25.5	26
	412 459 760	118.064 9	24.489 2	6.1	21.8	23
	413 694 950	118.064 8	24.483 8	8.3	231.8	235
	416 003 886	118.069 5	24.479 8	6.8	8.5	7
class 2	898 518 905	119.084 8	24.843 8	5.8	124.0	120
	200 051 220	119.084 5	24.843 9	5.1	91.1	88
class 3	412 704 150	118.107 2	24.402 0	16.7	316.5	314
	413 696 510	118.108 8	24.404 4	6.0	274.7	270
class 4	413 505 330	118.058 0	24.463 0	10.9	230.0	232
	23 660 006	118.060 6	24.463 5	8.4	264.6	259
class 5	371 602 000	119.069 6	24.773 5	8.1	225.6	226
	413 692 420	119.070 6	24.769 3	4.3	228.3	226
class 6	413 696 520	118.079 7	24.417 2	13.4	300.0	299
	413 700 450	118.078 6	24.413 2	6.7	129.9	130
class 7	412 592 180	118.866 6	24.722 5	5.6	222.4	224
	412 703 270	118.866 2	24.720 2	9.5	44.7	46
class8	412 469 870	118.074 0	24.505 8	8.8	18.0	17
	412 705 750	118.074 8	24.507 1	6.9	24.6	23

3.2 会遇类型划分

当会遇船舶相互位于各自正前方以相反或接近相反的方向相互逼近时, 通常是指船首向交叉在 6° 范围内构成对遇局面; 当船首向交叉大于 6° 且小于 112.5° 时, 构成交叉会遇; 若满足后船于前船正横大于 22.5° 范围内, 以较大速度赶超前船的情况, 则构成追越局面。其方位关系如图 4 所示。

根据图 4 方位关系, 结合会遇船舶相对方位关系和会遇船舶航向差可以实现不同会遇局面的识别。

3.3 会遇挖掘结果与分析

根据 DBSCAN 算法得出的潜在会遇船舶, 结合图 4 的会遇局面划分关系, 进行船舶会遇局面的识别与分析。本文选择追越、交叉和对遇 3 种典型局面进行展示, 结果如图 5 所示。

为避免碰撞发生, 图 5b 和图 5c 都有典型的避让行为发生。图 5b 中 A 船舶 (蓝色, 下同) 保持直航, B 船舶 (红色, 下同) 向右改向, 完成避让后回到原航向。图 5c 中 B 船舶为完成避让采取了向右改向, 待避让完成后回到原航向继续行驶。这也表示船舶严格遵守了《规则》中的向右改向避让他船的要求。

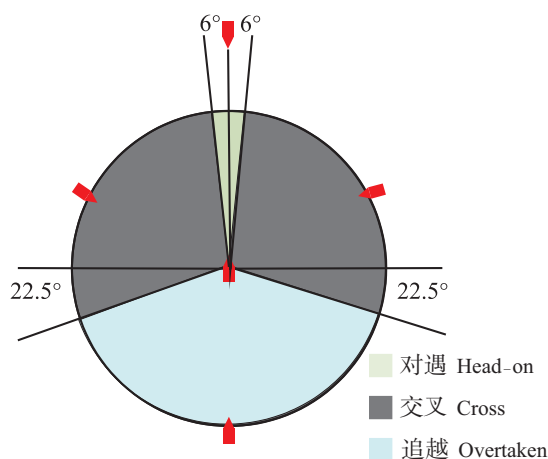
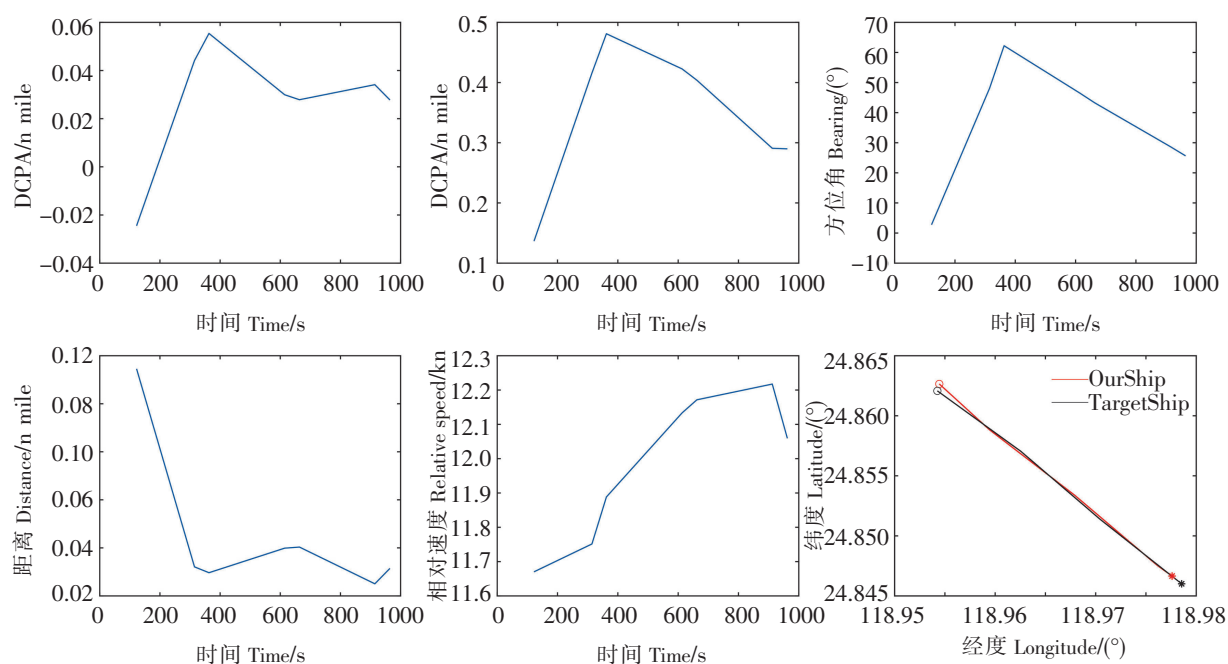
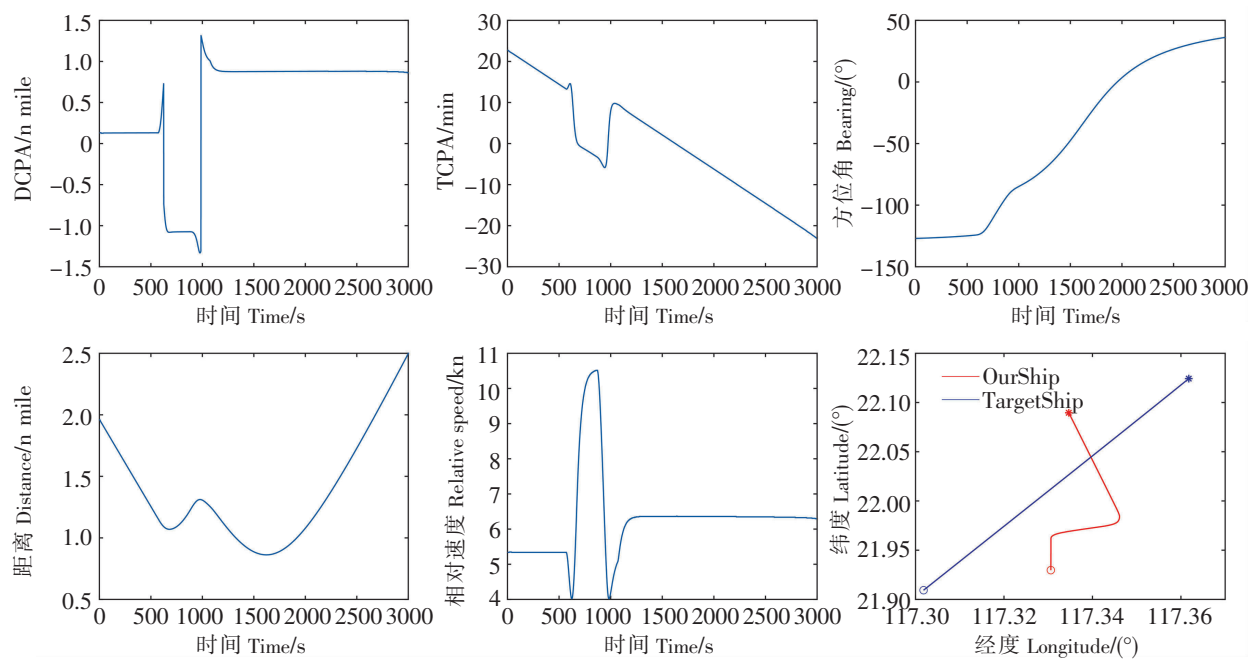


图 4 会遇局面方位关系

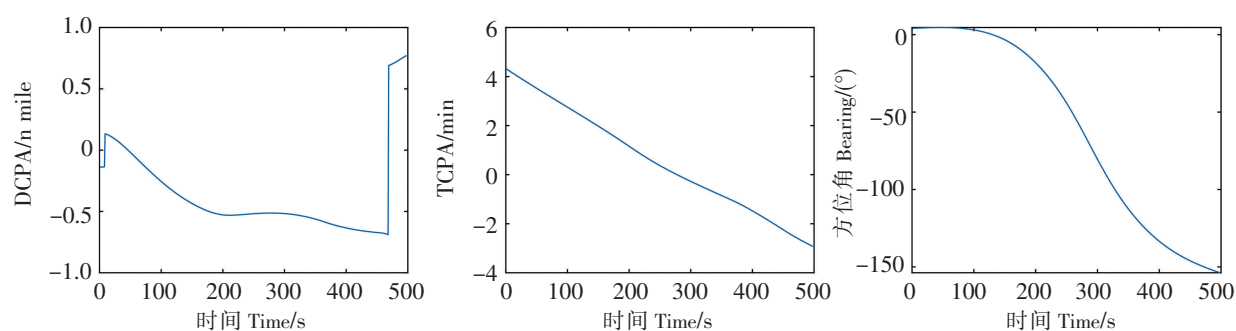
Fig.4 The positional relationship of the encounter situation

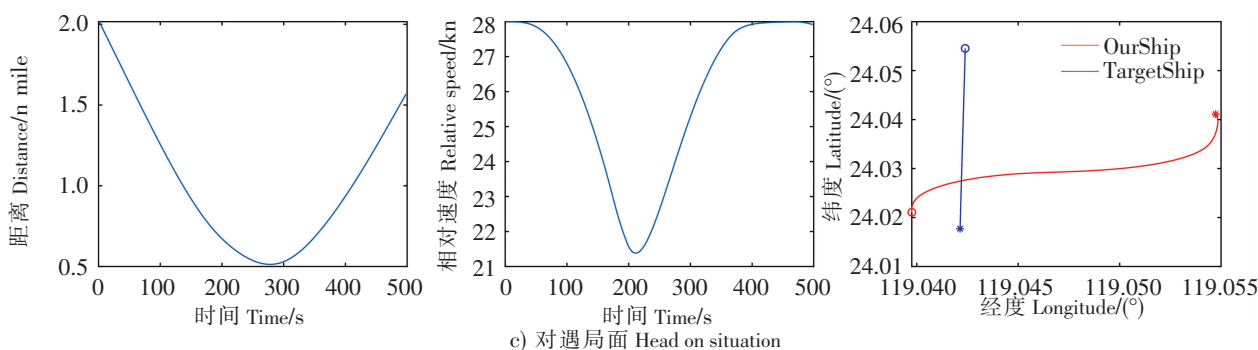


a) 追越局面 Overtake situation



b) 交叉局面 Cross situation





c) 对遇局面 Head on situation

图 5 船舶会遇局面

Fig.5 Ship encounter scenarios

4 结论

为保障船舶在海上的航行安全,提高船舶危险会遇的识别率,本文基于 AIS 数据,通过改进 DBSCAN 算法的相对距离矩阵,减小了 DBSCAN 算法的运算量。同时在数据不失真的前提下,合理处理 AIS 数据,减小了 CPA 算法的运算量,并且处理过程中保证了数据时空的一致性,减小了会遇的误判与漏判。本文以台湾海峡部分数据为基础,结合会遇局面方位关系,验证了算法的有效性。通过定量描述船舶会遇过程,可加深对会遇态势的整体感知,再现船舶会遇过程,为后续基于 AIS 的船舶避碰行为研究提供参考,也为自动避碰决策优化提供一种可行的验证方法。目前算法仅对两船会遇进行了研究,后续还需对多船会遇进行深入研究。

[参考文献]

- [1] 邓家佩. 基于 AIS 数据的群桥水域船舶碰撞危险区研究 [D]. 武汉: 武汉理工大学, 2019.
- [2] 潘家财, 姜青山, 邵哲平. 船舶会遇的时空数据挖掘算法及应用 [J]. 中国航海, 2010, 33(4): 57-60, 64.
- [3] 任亚磊. 基于 AIS 数据的船舶会遇特征研究 [D]. 武汉: 武汉理工大学, 2013.
- [4] 马杰, 刘琪, 张春玮, 等. 基于 AIS 的数据时空分析及船舶会遇态势提取方法 [J]. 中国安全科学学报, 2019, 29(5): 111-116.
- [5] 甄荣, RIVEIRO MARIA, 金永兴. 一种基于 DBSCAN 的船舶会遇实时识别方法 [J]. 上海海事大学学报, 2018, 39(1): 1-5.
- [6] 江龙晖, 郑中义, 齐乐. AIS 数据中船舶会遇信息的提取方法 [J]. 中国科技论文, 2017, 12(7): 802-805.
- [7] DU L, VALDEZ BANDA O A, GOERLANDT F, et al. Improving near miss detection in maritime traffic in the northern baltic sea from AIS data [J]. Journal of Marine Science and Engineering, 2021, 9(2): 80.
- [8] ZHANG W, GOERLANDT F, KUJALA P, et al. An advanced method for detecting possible near miss ship collisions from AIS data [J]. Ocean Engineering, 2016, 124: 141-156.
- [9] KIM D, HIRAYAMA K, OKIMOTO T. Distributed stochastic search algorithm for multi-ship encounter situations [J]. Journal of Navigation, 2017, 70(4): 699-718.
- [10] CHEN P F, HUANG Y M, PAPADIMITRIOU, et al. An improved time discretized non-linear velocity obstacle method for multi-ship encounter detection [J]. Ocean Engineering, 2020, 196: 106718.
- [11] 吴兆麟, 朱军. 海上交通工程 [M]. 大连: 大连海事大学出版社, 2004: 38-51.
- [12] 李威. 开阔水域船舶最小安全会遇距离模糊确定 [D]. 大连: 大连海事大学, 2016.
- [13] GAO M, SHI G-Y. Ship-handling behavior pattern recognition using AIS sub-trajectory clustering analysis based on the T-SNE and spectral clustering algorithms [J]. Ocean Engineering, 2020, 205: 106919.

(责任编辑 陈 敏 英文审校 周云龙)