

基于统计学理论的船舶轨迹异常识别

甄 荣, 邵哲平, 潘家财, 赵 强

(集美大学航海学院, 福建 厦门 361021)

[摘要] 为了准确识别异常航行轨迹的船舶, 以船舶 AIS 信息为数据源, 利用统计学中曲线拟合的最小二乘法对训练集船舶轨迹点进行拟合, 得到船舶典型航行轨迹的数学表达模型, 以此作为标准, 通过计算监控船舶轨迹点与典型轨迹间的距离是否大于典型轨迹 95% 置信区间的范围, 从而对轨迹异常的船舶进行识别. 实验结果表明, 该方法可以有效地识别轨迹异常船舶. 将该方法运用到监控系统中可以提高海上交通监控效率.

[关键词] 船舶; 航行轨迹; 统计学; AIS 信息; 监控

[中图分类号] U 675.7; TP 391

[文献标志码] A

A Study on the Identification of Abnormal Ship Trajectory Based on Statistic Theories

ZHEN Rong, SHAO Zhe-ping, PAN Jia-cai, ZHAO Qiang

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

Abstract: In order to identify the abnormal trajectory of a ship in the maritime traffic monitoring system automatically, sourced from shipborne AIS data, the curve fitting of statistic theories was used to fit the trajectory points of a training vessel, so as to obtain the mathematical model for the typical route of a ship. Based on this, ships with abnormal trajectories are identified through calculating whether the distance between validation data and typical route is wider than the 95% confidence interval of the typical route. The results show that abnormal ship trajectories can be identified efficiently this way. If this is applied to the vessel traffic surveillance system, the efficiency such surveillance will be increased significantly.

Key words: ship; trajectory; statistic theories; AIS Data; surveillance

0 引言

配备船舶自动识别系统 (Automatic Identification System, AIS) 的船舶能够自动发射船舶航行的静态和动态信息^[1], 这些 AIS 信息能够清晰地反映船舶实时航行状态和交通环境, 也为构建船舶典型航行行为模型提供了基础数据^[2]. 将船舶 AIS 历史信息作为训练集^[3]数据, 提取船舶典型运动模型, 从而对船舶的异常行为进行监控是当前研究的一个热点. 文献 [4] 将监控海域分为若干网格, 以高斯混合模型作为聚类模型, 用期望最大化模型估计得到每个网格内样本船舶航行轨迹的概率密度函数, 通过计算检测船舶的轨迹点是否符合对应网格的概率密度函数来识别船舶异常行为. 文献 [5]

[收稿日期] 2014-12-26

[修回日期] 2015-03-01

[基金项目] 福建省自然科学基金资助项目 (2012G0030)

[作者简介] 甄荣 (1990—), 男, 硕士生, 从事海上交通工程研究. 通信作者: 邵哲平 (1964—), 男, 教授, 船长, 博士, 从事交通信息工程及控制、航海技术研究, E-mail: zpsbao@jmu.edu.cn.

利用核密度估计得到船舶轨迹的概率密度函数来识别船舶异常行为. 文献 [4-5] 由于需要将监控区域分成网格, 当监控区域较大时, 网格的大小和数量的选取不容易确定, 从而影响识别效果. 本文运用统计学理论中曲线拟合的最小二乘法对 AIS 信息中包含的船舶航行轨迹点进行拟合, 以期得到船舶的典型航行轨迹数学模型, 通过计算监控船舶偏离典型航行轨迹的距离是否偏离典型轨迹的置信区间^[6]来对船舶异常行为进行识别, 使其不受监控区域大小和训练集数据多少的限制.

1 AIS 数据选取分析及预处理

1.1 实验数据选取分析

本文所采用的 AIS 数据由基于 ECDIS 和 AIS 的船舶综合信息服务系统^[7-9]提供. 船舶 AIS 数据包含信息可以分为静态信息、动态信息和与航次有关的信息, 其中静态信息包括: IMO 编码、船舶呼号、船名、船舶长度和宽度、船舶类型、定位天线在船上的位置等; 动态信息包括船位、国际协调时、对地航向、对地航速、航迹向、航行状态、转向率等; 与航次相关的信息包括: 船舶吃水、危险货物的类型、目的港和预计到达的时间、航行计划、简明安全信息等. AIS 信息中表征船舶轨迹点的信息为一系列离散的船位点, 为了对船舶航行轨迹点进行处理, 本文选取 AIS 动态信息中船舶的船位信息及信息发送时间、对地航向 (Course over Ground) 作为初始实验数据, 数据存储格式为 Ship (TIME, LONGITUDE, LATITUDE, COG).

1.2 数据预处理

由于 AIS 信息受到传输、解码等因素的影响, 其中有些信息存在错误和缺失, 因此有必要根据研究对象进行预处理, 主要预处理措施有^[10]: 删除动态信息表中 MMSI 为 0 的记录; 删除明显错误的 AIS 数据, 如经纬度超过合理值的数据 (如出现负值或经度 $> 180^\circ$ 或纬度 $> 90^\circ$).

为了验证本文方法的有效性, 将 AIS 数据分为训练集数据和校验集数据, 其中训练集数据用于提取典型航行轨迹数学模型, 校验集数据作为被识别、监控数据.

2 船舶轨迹异常识别过程

2.1 数学原理

2.1.1 船舶典型航行轨迹提取

在科学实验和工程实践中, 通常会涉及很多变量, 需要研究变量之间的关系, 很多时候变量之间的关系是不确定的, 需要用一个近似的函数来表示这种关系, 数据拟合就是根据变量的观测数据研究变量之间的近似函数关系, 用来帮助认识事物的内在规律和本质属性^[11]. 为了从大量的历史船舶 AIS 信息包含的航迹点中得到海上特定航行区域船舶航行的常态、典型轨迹表达模型, 本文采用统计学理论中曲线拟合的最小二乘法对历史船舶航行轨迹点进行拟合, 得到船舶典型航行轨迹的表达式. 曲线拟合的最小二乘法原理如下:

设船舶轨迹数据为 $\{(x_i, y_i), i = 0, 1, \dots, m\}$, x_i 和 y_i 分别表示训练集中相同船舶的经度和纬度, 为了表征监控区域船舶的典型航行轨迹, 设 $y_i = f(x_i)$, 采用最佳平方逼近法对训练集中船舶轨迹点进行逼近, 即在集合 $\varphi = \text{Span}\{\varphi_0, \varphi_1, \dots, \varphi_n\}$ 中找一组函数组合, 本文取 $\varphi = \text{span}\{1, x, x^2, \dots, x^n\}$ 使其到船舶轨迹点的距离误差的平方和最小, 即

$$\min_{S(x) \in \varphi} \sum_{i=1}^m \delta_i^2 = \min_{S(x) \in \varphi} \sum_{i=1}^m [S^*(x_i) - y_i]^2 = \min_{S(x) \in \varphi} \sum_{i=1}^m [S^*(x_i) - y_i]^2, (n < m). \quad (1)$$

其中: $S^*(x) = a_0^* \varphi_0 + a_1^* \varphi_1 + \dots + a_n^* \varphi_n$. 这样用最小二乘法拟合曲线的过程就是求解 $S^*(x)$ 的各项系数 $(a_0^*, a_1^*, \dots, a_n^*)$, 使其满足公式 (1), 最终转化为求多元函数极小值的问题, 具体求解过程参见文献 [11], 求得的典型船舶航行轨迹表达式并不是一条具体的轨迹, 而是最大程度地反映了监控海域船舶航行轨迹的公共特征数学模型.

2.1.2 船舶轨迹异常识别

在工业产品指标的合格性抽样检测中,经常考察抽样检测产品的测量指标是否落在合格样本均值的置信水平 95% 的置信区间内,来判断抽样检测产品的合格性,如果检测产品的指标位于置信区间之外,那么就将该产品识别为不合格产品,通常情况下以均值为中心,以正、负 2 倍标准差 (2σ) 为置信上、下限的置信区间所确定的置信水平为 95% [6].

根据以上原理,本文通过计算校验集中监控船舶航行轨迹到船舶典型航行轨迹的距离是否大于 95% 的置信区间,从而识别船舶异常轨迹,监控船舶距典型轨迹的距离计算方法为

$$D = |Ax_i + By_i + C| / \sqrt{(A^2 + B^2)},$$

(2)

其中 (x_i, y_i) 为监控船舶经、纬度坐标, A, B, C 为典型轨迹表达式的系数.

2.2 船舶轨迹异常识别原则

考虑到海上航行的船舶由于船舶自身吃水与可航水域水深的关系,在狭窄水道或进出港航道内,不同吃水船舶选择的横向航行位置不同.为了提高异常识别的准确性和有效性,在进行异常轨迹识别时,要保证训练集和校验集是具有相同特性船舶的轨迹数据.本文通过对特定船舶长度区间内的训练集船舶轨迹拟合得到的典型轨迹,来识别同样船长区间的校验集船舶轨迹的异常情况.

3 以厦门港青岛水道为例的船舶轨迹异常识别

采集 2014 年 5 月上半月进出厦门港青岛水道船舶 AIS 数据,为了保证校验集船舶轨迹与训练集船舶轨迹具有相同特性,选择船长为 100 ~ 200 m 的船舶轨迹点数据,其中前 1—10 日的数据作为训练集,用来提取典型航行轨迹,11—15 日的数据作为校验集,运用式 (2) 计算监控船舶位置点与典型轨迹的距离,识别偏离典型轨迹的船舶异常行为.

3.1 船舶典型航迹提取

厦门港青岛水道航行船舶的原始航迹展示如图 2 所示.该段航道是船舶进出厦门港的必经之地,航道内的船舶可分为进港和出港两大类,且船舶在航道内航行的横向位置符合《国际海上避碰规则》第九条中“靠右”航行的规则[12],因此假定训练集中的 100 ~ 200 m 船长船舶航行轨迹为常规的、正常的航行轨迹.

为了从训练集船舶轨迹数据中提取数学模型来表征所有船舶的航行轨迹分布,运用曲线拟合的最小二乘法原理对进出港船舶的轨迹点进行拟合, x 、 y 分别表示船舶轨迹点的经、纬度,得到的进、出港船舶典型轨迹表达式分别为:

$$\begin{cases} \text{Typical_route_in}: y = -0.989 * x + 141, \\ \text{Typical_route_out}: y = -0.969 * x + 139. \end{cases}$$

(3)

图 3 是经拟合得到的训练集船舶典型航行轨迹,由于选择航道为直线型,所以进、出港船舶典型轨迹是直线,且直线的斜率与青岛水道的走向是一致的,得到的典型轨迹最大程度地反映了船舶航行位置分布的公共特征.

3.2 进港船舶轨迹异常识别

当校验集数据船舶进入监控区域,首先根据船舶 AIS 信息中所包含的船舶对地航向(COG)来判断是进港还出港船舶,本文仅以进港船舶为例,运用式 (2) 计算监控船舶到进港船舶典型航行轨迹的距

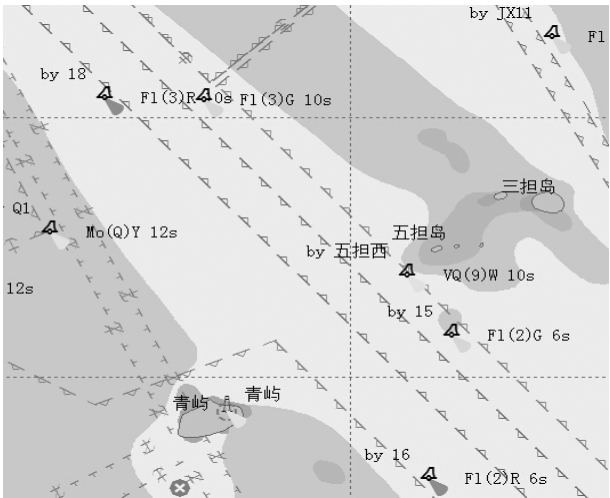


图 1 厦门港青岛水道
Fig.1 Xiamen Qingyu Channel

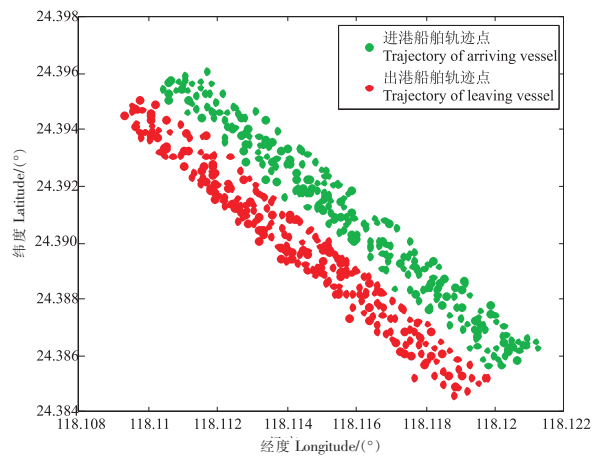


图 2 训练集船舶轨迹点
Fig.2 Training data

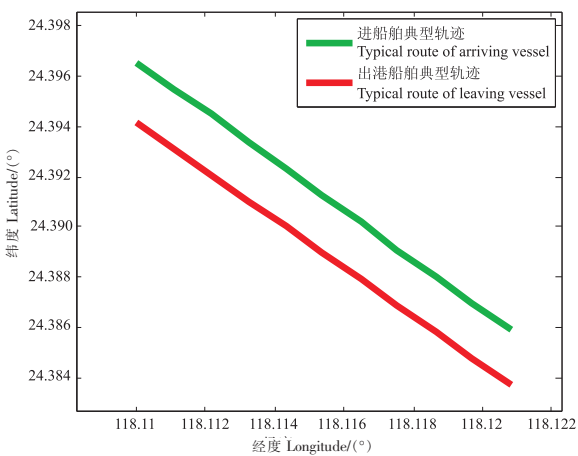
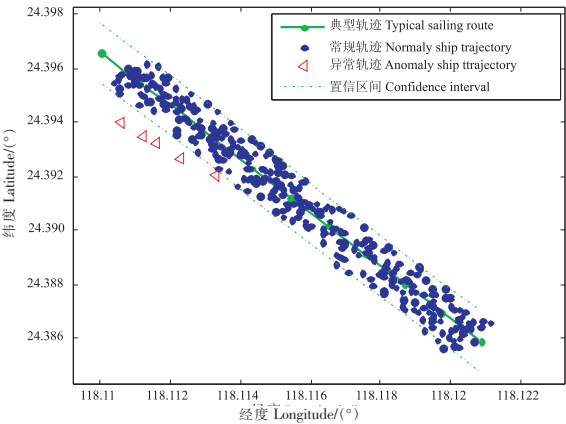


图 3 进出港船舶典型航行轨迹

Fig.3 Typical sailing route of arriving and leaving vessel

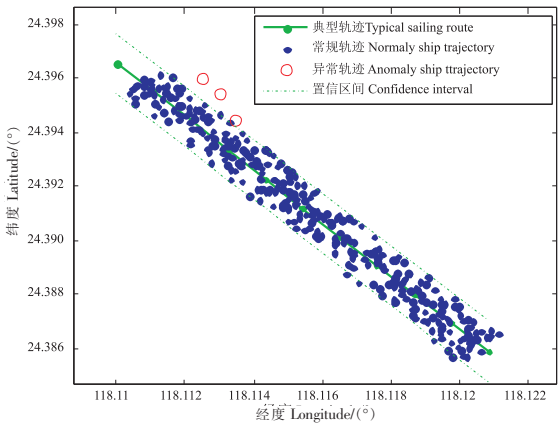
离,如果监控船舶在航行过程中到典型轨迹的距离大于典型轨迹 95% 的置信区间范围,那么计算机将该船舶自动识别为异常行为的船舶,从而提醒监控者给予该船舶更多的关注或实时的提醒。

如图 4 所示,实线为青屿水道进港船舶航行典型轨迹,两侧虚线是典型轨迹 95% 的置信区间边线,当有船舶驶入本海域时,计算船舶轨迹点到典型轨迹的距离,图 4 中实心点标识的船舶轨迹点位于典型轨迹置信区间边线内,属于常规航行轨迹。图 4a 中,用三角符号标识的船舶在航行过程到典型轨迹的距离大于左侧置信区间的范围,如果同时有船舶出港,这样的异常行为会妨碍出港船舶的安全航行;图 4b 中圆圈标识的船舶在航行的过程中偏离典型轨迹右侧置信区间的范围,偏离航道航行,通常航道以外的水域深度较浅,如果水深小于船舶吃水,可能会造成船舶搁浅事故。监控系统将轨迹异常的船舶标定以后,自动提示监控中心人员对这些轨道异常的船舶予以更多注意,从而减少监控的盲目性。监控人员可以对轨迹异常船舶予以警告和提示,更好地保障海上交通安全。



a 偏向典型轨迹左侧

Portside anomaly trajectory of typical route



b 偏向典型轨迹右侧

Starboardside anomaly trajectory of typical route

图 4 青屿水道进港船舶轨迹异常识别结果

Fig.4 Arriving vessel trajectory anomaly identification result in qingyu channel

4 结论

海上交通监控在确保海上交通安全、提高海上交通效率中发挥着重要的作用,本文利用统计学理论中曲线拟合的最小二乘法对历史 AIS 信息中训练集船舶轨迹点进行拟合,得到监控海域船舶的典型航迹,以此作为参考,对监控海域偏离典型航迹 95% 置信区间的异常行为的船舶进行有效地识别,可以降低

低监控人员的工作强度. 本文初步实现了大型海港进出港航道内船舶偏离典型航行的船舶异常行为的有效识别,对穿越主航道航行、船舶速度过高或过低等异常行为的识别将是今后的研究方向.

[参 考 文 献]

- [1] 孙文力, 孙文强. 船载自动识别系统[M]. 大连: 大连海事大学出版社, 2004.
- [2] 吴兆麟, 朱军. 海上交通工程[M]. 2 版. 大连: 大连海事大学出版社, 2004.
- [3] CHRISTOPHER M BISHOP. Pattern Recognition and Machine Learning[M]. New York, USA: Springer Science Business Media, 2006.
- [4] RIKARD LAXHAMMAR. Anomaly detection for sea surveillance[C]//The 11th International Conference on Information Fusion. Cologne, Germany: IEEE, 2008.
- [5] RISTIC B, SCALA B L A, MORELANDE M, et al. Statistic Analysis of Motion Patterns in AIS Data: Anomaly Detection and Motion Prediction[C]//The 11th International Conference on Information Fusion. Cologne, Germany: IEEE, 2008.
- [6] 盛骤, 谢式千, 潘承毅. 概率论与数理统计[M]. 北京: 高等教育出版社, 2008.
- [7] 邵哲平, 孙腾达, 潘家财, 等. 基 ECDIS 和 AIS 的船舶综合信息服务系统的开发[J]. 中国航海, 2007(2): 30-33.
- [8] SHAO Z P, SUN T D, PAN J C, et al. Vessel information service system based on ECDIS and AIS[C]//Proceedings of ICTE. [S. L.], America: ASCE, 2007: 1678-1683.
- [9] JI XIANBIAO, SHAO ZHEPING, PAN JIACAI. A New AIS-based Way to Conduct OLAP of Maritime Traffic Flow [C] //ASCE, Proceeding of ICTE 2009, [S. L.], America: ASCE, 2009.
- [10] 唐存宝, 邵哲平, 潘家才, 等. 基于 AIS 的船舶航迹分布算法 [J]. 集美大学学报: 自然科学版, 2012, 17(2): 110-111.
- [11] 何正风. MATLAB 概率与数理统计 [M]. 北京: 机械工业出版社, 2012.
- [12] 吴兆麟, 赵月林. 船舶值班与避碰 [M]. 3 版. 大连: 大连海事大学出版社, 2009.

(责任编辑 陈 敏 英文审校 周云龙)