

船舶低效航行状态的识别与分析方法

张健豪, 韩懿

(中远海运科技股份有限公司, 上海 200135)

摘要: 经济全球化给物流行业带来了远大的前景, 集装箱运输以其低廉的价格和巨大的运载量成为国际贸易的最重要的运输方式。本文首次运用统计学方法中的 fused lasso 算法和变点分析方法对集装箱船舶航速数据进行分析, 结合船舶自动识别系统数据的特点和集装箱航行特征, 对有效航行时间进行分析, 完成航行模式的识别, 掌握船舶在港口拥堵状况。

关键词: 统计机器学习; 航行行为识别; 港口拥堵状况

中图分类号: U66 **文献标识码:** A **文章编号:** 1006—7973 (2022) 08—0095—03

1 引言

自 2020 年以来, 全球新冠疫情蔓延, 港口拥堵已成为物流供应链中的瓶颈, 给班轮运输的稳定性造成了很大的影响。美国是感染人数较多的国家, 由于码头工人和卡车司机大量感染, 造成人员短缺, 长滩港和洛杉矶港出现了长期拥堵, 国际海运物流供应链受到严重影响。船舶在港拥堵滞留问题, 直接增加了航运企业的经营成本, 以及影响全球物流供应链的正常运行。据统计, 2021 年 90% 停靠美国洛杉矶港装卸作业的船舶直接向锚泊等待, 平均锚泊时间达 7.8 天。据美国 CNN 新闻报告, 洛杉矶港和长滩港在 2021 年的漂航船舶数量达到历史峰值。

正常情况下港口拥堵表现为船舶锚泊时间长, 但洛杉矶港和长滩港的情况无法用锚泊时间来衡量, 因为锚地有限, 引发大量的船舶漂航, 这对人力成本和能源消耗是一种巨大的浪费。目前业界对港口拥堵期船舶无效航行的研究较少, 无法直接获得准确的无效航行时间。我们现在提出一种基于船舶自动识别系统 (Automatic Identification System, 以下简称“AIS”) 中船舶航行模式的识别方法。该方法是通过设定目的港考察相关集装箱船舶的航行速度和里程的变化, 通过分段计算速度和里程找到并标识无效航线开始的时间和位置, 判断船舶的航行状态进而计算在港船舶等待时间来评估港口的拥堵状态, 用于航运经营人合理调配运输航线、港口生产调度和行业管理部门决策参考。本文通过一系列统计变点分析方法处理和分析 AIS 数据, 并通过真实案例和历史数据验证了提出方法的实用性和有效性。

2 研究方法

2.1 主要原理

目前 AIS 系统中对船舶航行状态规定: 0 代表“航行”, 1 代表“锚泊”, 5 代表“系泊”。但 0 状态无法识别漂航, 因此仅看这个字段会产生误判, 我们无法区分有效航行和无效、低效航行。

我们通过利用 AIS 动态数据, 基于一些统计学习的相关算法, 并结合业务逻辑, 提出一种新颖的船舶航行模式识别。我们提出的航行行为识别方法来源于统计信号识别, 此方法可以有效地将航程分割成若干段, 为后续的数据分析提供指导意义。

2.2 实施步骤

现以集装箱班轮运输为例, 通过数据采集和预处理等五个步骤来实施。

(1) 数据采集和预处理。识别航线过程中需要集装箱船的 AIS 静态和动态信息、集装箱班轮挂港情况、港口信息等, 具体来说包括船舶的水上移动通信业务标识码 (Maritime Mobile Service Identify, 以下简称“MMSI”)、船型、每个航段的起始和结束时间及起终点港口、港口所属的国家和地区、港口经纬度。获得上述数据后, 先对 AIS 数据进行预处理, 去掉 AIS 有问题 (如缺失过多)、修船、起始港和终点港有缺失或港口为未知的数据以及起点港与终点港为相同港口的数据, 并将起始港和终点港分别匹配到相应的国家和地区。由于 AIS 数据的采集不均匀, 我们需要对任何两个港口之间的 AIS 数据进行降采样处理, 通过降采样, 数据量将由原来的上万条变成上千条, 大大地减少了计算成本, 提高了运行速度。

(2) 基于机器学习算法识别船舶航行行为。基于上述处理过的 AIS 动态数据, 我们使用 fused lasso 算法^[1,3], 以及变点分析 (change point analysis) 方法^[2]对集

装箱船舶的航行行为进行数据挖掘。

(3) 结合业务逻辑对识别出的航线进行进一步处理。由于 fused lasso 算法和变点分析方法只是从统计学的角度对航速进行处理, 我们同样也需要结合业务逻辑对获得的分段航速进行解释。在这里我们规定接近 20 节(海里每小时)的航段代表高速正常航向, 接近 0 节的航段代表无效航行(漂航)。

(4) 航线相关信息提取及分类。为了避免速度分段过程中出现由于航速缺失造成找到的航段不匹配的问题, 数据需要进一步地清洗。我们添加航速的插值数据, 对缺失的航速数据进行补全; 根据航线经过的港口信息, 将航线匹配到相应的国家和地区; 找出航段的起始时间和终止时间, 确定该航线运营的相关年份时间信息; 根据目的港对航行行为进行分类, 包括近期拥堵港口以及正常港口。

(5) 航线航行行为识别及监控。对于设定的任意两个港口, 我们可以找到所有航行于这两个港口船舶的航行行为, 根据已经定义好的航线进行航速统计和航行行为分析。

3 主要创新点

(1) 创新识别船舶航行状态。运用统计学方法, 结合船舶自动识别系统数据特点和集装箱航行特征, 完成船舶航行模式的识别。

(2) 创新航行行为标准, 为航行监控和港口拥堵分析提供了支撑。同时保留了各航线的船型信息, 实现对任意一个目的港港口的任意一个时间段的船舶航行行为, 即航线、船型的动态查询。

(3) 计算出真实航线时间和无效航行时间。

4 真实数据分析

下面我们以 MMSI 为 373233000 的集装箱船为例, 对提出方法的可行性和有效性进行验证。

4.1 数据采集和预处理

首先从数据库中取出 MMSI 为 373233000 的集装箱船的 AIS 历史动态数据。清洗后的数据如图 1 所示, 其中图中的每一点代表单位小时内的平均速度, 通过计算整点时间内实际航行距离, 以及整点时间内的实际航行时间的比值得到。

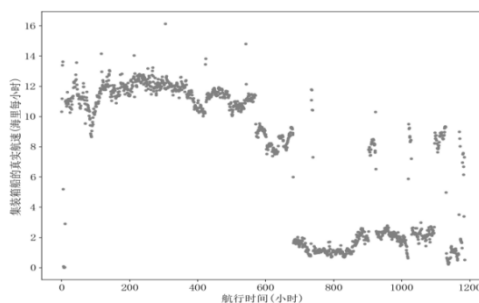


图 1 MMSI 为 373233000 的集装箱船数据预处理结果

4.2 基于统计学习算法识别航行行为

根据清洗后数据, 我们做出船舶的历史航速图, 并使用 fused lasso 算法和变点分析方法寻找历史航速的多个突变节点。两种方法的结果如图 2 所示, 可以看出识别结果的差别不大, 两种方法都可以准确地刻画航速变化趋势。为了方便阐述, 下文仅以变点分析方法的结果进行解释和分析。

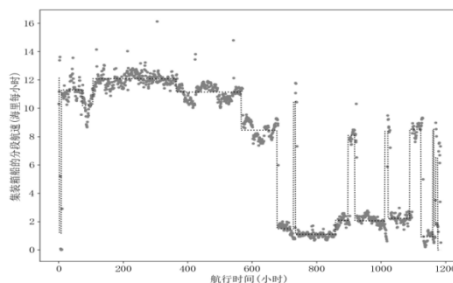


图 2 MMSI 为 373233000 的航速使用变点分析方法分段结果

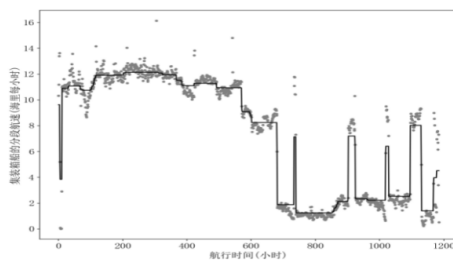


图 3 MMSI 为 373233000 的航速使用 fused lasso 分段结果

4.3 结合业务逻辑对识别出的航行行为进行进一步处理

结合业务逻辑, 对上述航行过程进行时间信息匹配, 可以识别出大致四个航行状态, 它们分别是: 低速启航段, 高速平稳段, 减速突变段, 无效漂航段。其中最复杂并且最需要的是无效漂航段。通过分析, MMSI 为 373233000 的船舶真实航行时间为 496 小时。

由图 2 可以看出, 该集装箱船在开始航行的第 600 多小时开始航速小于 2 节, 且持续时间长达约 52 小时, 此时的航行距离已经超过港间距的上限, 所以可以判定

该船有无效航行状态也就是漂航状态。一般基于数据观测的经验,我们可以设定无效航行状态为:航速小于2节持续12小时并且航行距离超过港间距的70%。一旦识别出漂航的开始的时间节点,我们可以计算开始漂航到开始靠泊的时间,并将此段时间作为漂航的持续时间。

4.4 航线相关信息提取及分类

提取目的港为长滩港和洛杉矶港在2021年9月到2022年1月内的所有集装箱船,根据之前的方法做类似的数据分析。最终结果显示有漂航状态的集装箱船舶大约占有所有集装箱船的20%,说明了港口拥堵非常严重。

4.5 集装箱班轮航线行为识别及监控

根据已经识别出的航行行为可以对经过任意两港的集装箱班轮航线进行航行行为识别,进而可以进行港口拥堵分析。

5 港口拥堵中的应用

目前AIS数据中缺少判断漂航的字段,不能自动识别港口真实拥堵情况。为了更加有效地掌握每个目的港的拥堵状态,我们以2021年目的港为洛杉矶港和长滩港的所有集装箱船舶为例,判断出漂航的船舶,并记录漂航船舶每天累计的漂航时间,更加具体地刻画出洛杉矶港和长滩港的拥堵状况。

2021年11月16日,为了减少洛杉矶港和长滩港附近锚泊的船舶,改善空气质量,提高港口的安全性。太平洋海事协会、太平洋商船协会和南加州海事交易所制定一种新的集装箱船排队流程,该程序让船只在外海等待,而不是在港口附近的锚地和游荡区等待。

由于现有通过港口等泊时间评价拥堵情况不准确,比如洛杉矶港和长滩港拥堵前期拥堵,锚地狭小,已有部分船舶产生漂航,洛杉矶港和长滩港新政后,港口排队禁止锚泊进一步加剧了漂航,大量船舶开始漂航,因此无法获得漂航的时间和里程。

以2021年的AIS数据为例,在图4中我们可以发现在洛杉矶港和长滩港2021年11月16日出台新规则之前,每天累计漂航集装箱船舶的数量长期处于高位,因此代表港口的严重拥堵。随着新规的实施,漂航船舶数量得到极大地减少。同样的方法可以应用在全球所有的港口中,从而可以实时掌握港口的拥堵状态。

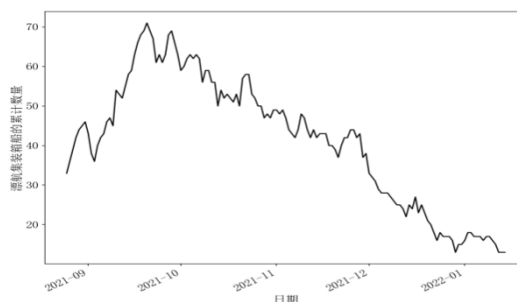


图4 2021年9月到2022年1月间目的港为长滩港和洛杉矶港的

漂航集装箱船舶数量走势图

6 总结与展望

随着海上交通智能化,船舶配备了大量先进的信息技术、通讯技术和网络技术,从而产生海量的数据。对航运大数据,通过结合实际不断挖掘数据的内在价值,让数据在科学管理、经营决策等方面发挥应有的作用,通过不断创新数据运用,为航运业高质量发展提供支持保障。本文基于统计机器学习来识别船舶低效航行状态行为新方法,用于判断港口的拥堵情况,为港航企业经营管理决策参考。

参考文献:

- [1] Arnold, T.B., & Tibshirani, R.J. (2014). Efficient Implementations of the Generalized Lasso Dual Path Algorithm. *Journal of Computational and Graphical Statistics*, 25, 1 – 27.
- [2] Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500), 1590–1598.
- [3] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005), "Sparsity and smoothness via the fused lasso", *Journal of the Royal Statistics Society: Series B* 67(1), 91–108.