

一种基于聚类分析的船舶航行工况划分方法

谭笑, 关文渊, 李晗, 李永杰, 薛晨

中国船舶工业系统工程研究院, 北京 100036

摘要 结合船舶航行工况与主机燃油消耗特性的关系, 考虑吃水和相对风速影响因素, 采用 K-means 聚类分析方法, 实现油耗影响因素的不同航行工况的划分。以某超大型油轮为例, 基于船舶航行装载与外部环境天气状态数据, 实现了航行工况分类分析以及各工况下影响因素参数区间的确定, 为船舶主机燃油消耗分工况匹配模型构建提供了更加精细化的分析基础。

关键词 船舶; 低速柴油机; 燃油消耗; 工况划分; 聚类分析

船舶运营成本包括燃料、人员工资、润滑油物料、维修保养等费用, 其中耗用燃料的费用占比非常大。每条船的燃油消耗来自主机、电站、锅炉等设备, 主要消耗是主机油耗, 其燃油消耗通常占全船燃油消耗的 90% 以上^[1-2], 极大地影响着航运企业的船舶运营成本和海洋大气环境。在目前航运市场环境中, 如何有效地进行船舶航行节能性优化和环保性提升, 成为各航运企业用户极为关心并亟待解决的问题。

为了降低主机燃油消耗, 通常进行主机建模与优化。基于船舶的实际运行数据, 建立主机燃油消耗优化的模型^[3], 调整船舶操纵或航速, 是控制主机燃油消耗成本的常用方式^[4-5]。由于在不同的工况下, 主机最低燃油消耗率不同, 为了获得更加准

确的模型和更加有效的优化结果, 航行过程中不同类型工况的合理划分十分必要。

在以往的船舶航行工况划分方法中, 几乎都是基于相关人员的经验, 通过相对随意的方法进行划分的。传统的船舶航行工况划分方法, 具有两大痛点: 数据获取维度少、数量小、不够准确, 船舶与环境全面准确数据的采集和存储方面存在空白, 缺少必要的资源; 缺乏有效的手段定量分析航行工况, 没有从大数据角度进行分析挖掘与范围评估, 无法根据船舶本身状态数据和环境数据的变化精确地作出定量化分析及划分。

聚类属于无监督的机器学习方式, 是根据未知标签样本的数据集内部数据特征, 将数据集划分为多个不同的类^[6], 适用于对船舶航行工况的划分。

收稿日期: 2018-10-15; 修回日期: 2019-01-25

基金项目: 智能船舶 1.0 研发专项(工信部联装函[2016]544 号)

作者简介: 谭笑, 工程师, 研究方向为通用工业大数据分析, 电子信箱: tanxiao2016cyn@163.com

引用格式: 谭笑, 关文渊, 李晗, 等. 一种基于聚类分析的船舶航行工况划分方法[J]. 科技导报, 2020, 38(21): 91-95; doi: 10.3981/j.issn.1000-7857.2020.21.011

在船舶领域现有相关技术中,大多采用人为经验划分,容易产生错误和遗漏^[7-8]。本研究提出一种基于聚类分析算法的船舶航行工况划分方法。

1 聚类算法原理

影响主机燃油消耗的原因,除主机自身的因素(燃油供给系统,扫气、换气系统,燃烧室,定时控制系统等)外,还包括船舶吃水(装载量)、船舶污底、气象条件(风浪情况)、船舶拖拽、窄航道或浅水航行等维度。结合船舶主机推进燃油消耗的机理,在船型不变的情况下,船舶载重和天气情况的外界条件同时影响着船舶阻力,决定了推进系统的运行功率,不同工况下的主机运行功率和燃油单元有密切关系。吃水的大小不仅取决于船舶和船载所有物品重量,而且还取决于船舶所处水的密度,相比载重数据融合了更多因素,并由于具有成熟的传感技术,容易测量。天气条件数据通过风速风向仪获得。结合船舶正常航行实际环境,以船上装载传感器获取运行分析数据为前提,选择船舶吃水、风的情况(风向和风速,由此可得出相对风速)这2个参数作为工况划分的主要影响因素维度,其数据样本皆没有类别标签 y ,只有特征 x 。

船舶吃水(draught)一般指船舶浸在水里的深度,是指船舶的底部至船体与水面相连处的垂直距离。船舶吃水情况直接影响船舶运行过程中的稳定性和结构安全。

船舶航行在有风海面上所测得的风速,由船风和风速合成而得。由于在目前绝大多数的船舶监测系统中,关于船舶气象环境相关的参数,仅对相对风速有较高集成度的测量与采集,对其他参数例如浪高、涌、海流等信息,均无法获得直接有效的数据,因此取相对风速数据作为船舶气象环境相关参数的代表进行分析。

船舶航行时的柴油机燃油消耗,同时受船舶自身条件和外界条件影响。因此,采用船舶吃水和相对风速,从船舶自身载货状态和外部环境2个维度,通过数据驱动的手段,建立船舶主机运行工况准确划分后的主机油耗工况匹配模型。在船舶运

营过程中,根据吃水和相对风速的数据分布情况,判断柴油机油耗工况的属性,为更精细化的船舶能效优化与辅助决策支持功能奠定基础。

船舶航行的工况本来没有分类。K-means算法是典型的基于距离的聚类算法,它采用距离作为相似性的评价指标,认为簇是由距离靠近的对象组成的,因此把得到紧凑且独立的簇作为最终目标^[9]。在聚类问题中,训练样本为 $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$,每个 $x^{(i)} \in R^n$,没有标签 y 。K-means算法将样本聚类成 k 个簇(cluster),具体算法描述如下。

1) 随机选取 k 个聚类质心点(cluster centroids)为 $\mu_1, \mu_2, \dots, \mu_k \in R^n$ 。

2) 重复下面过程直到收敛。

对于每一个样例 i ,计算其应该属于的类

$$c^{(i)} = \arg \min_j \|x^{(i)} - \mu_j\|^2 \quad (1)$$

对于每一个类 j ,重新计算该类的质心

$$\mu_j = \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}} \quad (2)$$

式中, k 是事先给定的聚类数, $c^{(i)}$ 代表样例 i 与 k 个类中距离最近的那个类, $c^{(i)}$ 的值是 $1 \sim k$ 中的一个。

质心 μ_j 代表对属于同一个类的样本中心点的猜测,对船舶航行工况模型来说就是要将所有的吃水和相对风速的向量聚成 k 个工况类型,首先随机选取 k 个吃水和相对风速的向量作为 k 个工况的质心,然后第一步对于每一个向量计算其到 k 个质心中每一个的距离,然后选取距离最近的那个工况作为 $c^{(i)}$,这样经过第1)步,每一个向量都有了所属的工况;第2)步对于每一个工况,重新计算它的质心 μ_j (对里面所有的向量坐标求平均)。重复迭代以上2步直到质心不变或者变化很小。

采用K-Means聚类算法的优点是:(1) 算法快速、简单;(2) 对大数据集有较高效率并且有可伸缩性;(3) 时间复杂度近于线性,而且适合挖掘大规模数据集。

本问题中,数据输入样本 x 为平均吃水 $D(i)$ 与相对风速 $s_R(i)$ 组成的向量数据 $[D(i), s_R(i)]$,数据样本本身不拥有任何分类意义的标签 y ,利用K-

means 算法就船舶航行工况的划分问题进行无监督学习分析和应用。

2 整体方案步骤

1) 数据获取。对于船舶的一段航段航行, 提取其航行过程中采集存储的多项参数数据, 包括主机运行相关数据、船舶状态数据、船舶设备数据及气象数据等。本方法将主要使用船舶状态数据中的吃水数据(包括船舶左舳吃水、右舳吃水)和气象数据中的风向、风速数据。以上数据通过船上安装各类传感器采集, 由机舱监测系统及其他配备的集成化系统存储和提供。数据进行如下预处理。

(1) 船舶平均吃水计算。将船舶左舳吃水 d_L 、右舳吃水 d_R 通过式(3)计算得到船舶平均吃水 D 。

$$D(i) = \frac{d_L(i) + d_R(i)}{2} \quad (3)$$

(2) 相对风速计算: 将风向 d_w 、风速 s_A 数据进行预处理, 通过式(4)计算得到相对于船首向的相对风速 s_R 数据:

$$s_R(i) = -s_A(i) \cos d_w(i) \quad (4)$$

式中, 正负可以表示风向与船首的顺逆关系, 正表示顺风, 负表示逆风。

每一组船舶平均吃水和相对风速数据向量 $[D(i), s_R(i)]$, 组成一个样本的数据。

2) 数据清洗。(1) 提取船舶稳定航行状态下主机运行相关样本数据。(2) 样本数据奇异值剔除。由于传感器异常、船舶操纵等原因带来的奇异值, 并非由工况的差异性产生, 对于航行工况聚类过程计算不具有关联意义, 通常采用设置阈值范围的方法进行剔除。

3) 假设航行工况初始聚类数 k , 进行 K-means 聚类算法计算。首先, 聚类的初始类别数量 k , 需要结合业务的理解与分析来确定。其次, 在 K-means 聚类算法的具体应用中, 采用欧氏距离(Euclidean distance)计算方法描述数据个体之间的相似度。这一步可以得到将所有航行工况数据样本进行初始聚类数时的聚类分析结果。

4) 优化聚类数 k , 重新进行聚类, 确定船舶航

行工况划分。根据第3)步得到的分析结果, 进行聚类数 k 的优化, 直到聚类分析结果效果最佳, 使工况的划分能够满足使船舶装载状态吃水明显区分(不能将不同载重下的吃水数据判断在同一个样本中心点质心的分类以内)、环境因素的相对风速划分有合理涵义。将优化后的聚类数 k 代入 K-means 聚类算法得到聚类分析结果^[10]。通过每个航行工况簇中的实船数据, 经过统计计算得到一个航行工况分类的船舶吃水和相对风速 2 个参数的范围区间。

聚类数 k 的优化调整标准, 需要满足一个聚类中心中的样本能够完全区分出压载和满载的区别, 即一个聚类中心中的样本不应该既有压载又有满载的数据, 否则将不符合实际航行使用的需求。方案的简要原理流程如图 1 所示。

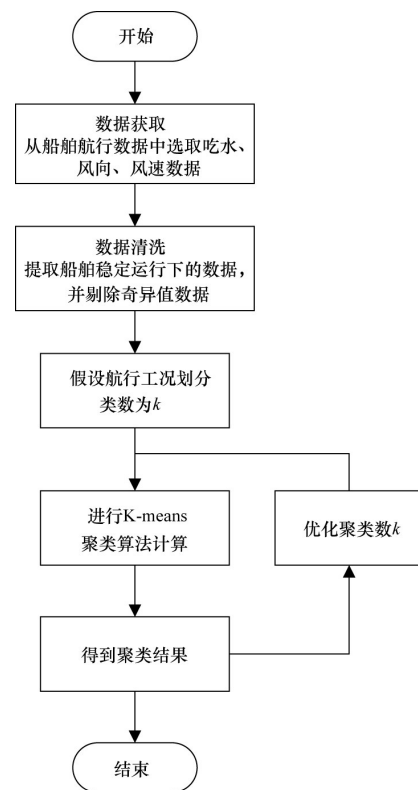


图1 K-means 聚类分析的船舶航行工况划分方案简要原理

3 实船应用结果

针对某超大型油轮 VLCC 在某 2 个月的航段应

用上述方案。首先,对于该船在此航段的航行,提取它的实际航行过程中采集存储的多项参数时间序列数据,整理并计算得到船舶平均吃水和相对风速数据向量 $[D(i), s_r(i)]$,组成样本数据集。进行数据清洗时,通过对主机运行相关数据分析计算,提取船舶稳定航行状态下的样本数据。并设置阈值范围,剔除船舶平均吃水和相对风速的奇异值数据(根据船型和船舶下水实际情况可以确定,对于该VLCC的吃水数值的正常阈值范围为0~25 m;根据航行经验,相对风速数值的正常范围为-40~70 m/s)。对2个考察维度进行统计分析。对吃水分布(图2)进行观察,结合大部分油轮都是单程运载原油运行,去程大都为压载水压载保证航行状态,返程为满载航行状态,因此吃水情况分压载和满载2类状态。对相对风速分布(图3)进行观察,特征并不明显,通常根据航行中风的顺逆情况和风力大小,可将相对风速状态初步分成顺风、小逆风和大逆风3类^[11-12]。

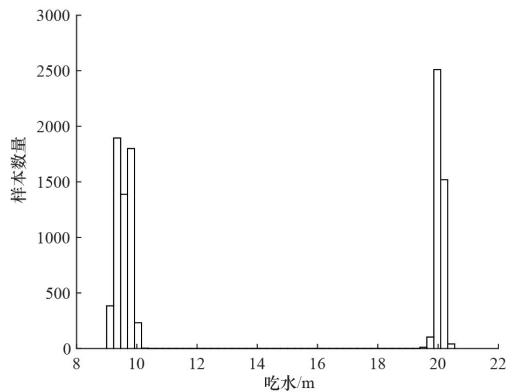


图2 船舶吃水数据分布

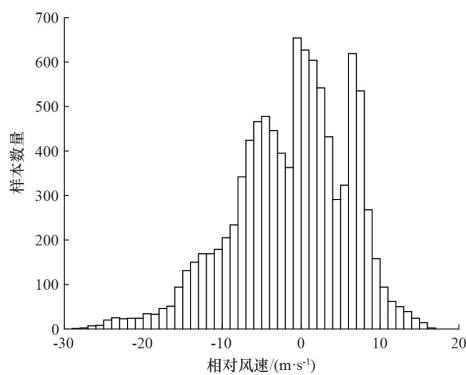


图3 相对风速数据分布

因此,首先假设航行工况聚类数 $k=2 \times 3=6$,代入K-means聚类算法计算,得到分析结果如图4(a)所示。可以看出,存在一个聚类中心中的样本并没有区分出压载和满载的区别,不能够满足实际的工况应用要求,为了使分类结果更加可靠,设置新的聚类数为 $k=7$,重新进行聚类分析计算,得到聚类分析结果如图4(b)所示。

最终得出该超大型油轮几种航行工况划分结果以及每个工况类型的吃水和相对风速2个参数量化范围区间,结果如表1所示。

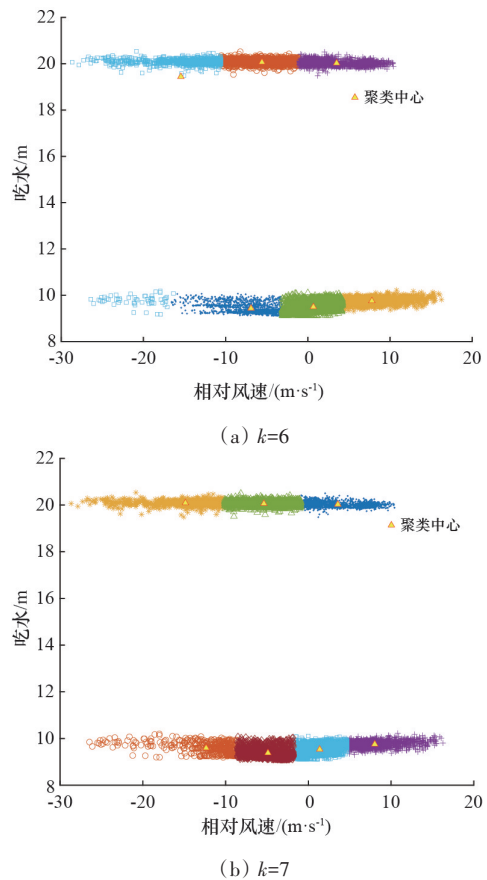


图4 聚类结果

表1 工况类型吃水和相对风速参数量化范围区间

编号	航行工况类型	相对风速范围 $/(m \cdot s^{-1})$	吃水范围/m
1	压载、顺向和风	$[4.70, +\infty)$	(0, 15)
2	压载、轻风	$[-1.76, 4.70)$	(0, 15)
3	压载、逆向和风	$[8.60, -1.76)$	(0, 15)
4	压载、逆向强风	$(-\infty, -8.60)$	(0, 15)
5	满载、顺风	$[-0.90, +\infty)$	[15, 25)
6	满载、逆向和风	$[-10.11, -0.90)$	[15, 25)
7	满载、逆向强风	$(-\infty, -10.11)$	[15, 25)

4 结论

提供了一种基于聚类分析方法的船舶航行工况划分方法,解决了数据资源获取维度少、数量小、不够准确,以及缺乏有效的手段定量挖掘分析航行工况的2个痛点,通过对实船航段实际航行数据进行分析处理,航行工况划分的结果在细化程度上有所提高。进而得到船舶典型航行工况基于吃水和相对风速因素的划分结果。本方法存在对噪声和离群点敏感,结果不一定是全局最优只能保证局部最优的局限性。然而其原理简单,易在智能船舶相关应用系统中实现,在实际场景中能够真实反映船舶航行时主机面临的不同环境状况,符合船舶运行实际过程,可为主机燃油消耗的建模和优化提供基础。

参考文献(References)

- [1] 刘啸波,任慧龙,李陈峰. 世界及我国船用低速柴油机行业发展研究[J]. 船舶工程, 2013(1): 1-4.
- [2] 李思锦. 船舶动力系统的设计与研究[D]. 大连: 大连理工大学, 2012.
- [3] 周春斌,袁成清,张彦. 船舶主机油耗模型发展现状及展望[J]. 柴油机, 2017(6): 12-15.
- [4] 赵乾博. 考虑油耗及经济性的营运船舶定航线航速优化[D]. 哈尔滨: 哈尔滨工程大学, 2017.
- [5] 苏一,张倩墨. 基于能效管理的船舶航速系统优化设计[J]. 上海船舶运输科学研究所学报, 2014, 37(2): 43-49.
- [6] 李卫军. K-means 聚类算法的研究综述[J]. 现代计算机(专业版), 2014(23): 31-32.
- [7] 刘伊凡,黄连忠,孙培廷,等. 实船数据环境下船舶主柴油机性能评估方法[J]. 内燃机学报, 2018(2): 182-183.
- [8] 施浩,黄亦翔,刘成良. 基于多工况自动聚类的飞机发动机健康状态评估[J]. 机电一体化, 2016, 22(11): 39-43.
- [9] 吴凤慧,成颖,郑彦宁,等. K-means 算法研究综述[J]. 现代图书情报技术, 2011, 27(5): 28-35.
- [10] 任景彪,尹绍宏. 一种有效的K-means 聚类初始中心选取方法[J]. 计算机与现代化, 2010(7): 84-86.
- [11] 马静,高翔,李益楠,等. 考虑风速随机特征的多工况电力系统稳定性分析[J]. 电力自动化设备, 2016, 36(8): 26-32.
- [12] 李美菁,吴秀恒. 船舶在风浪流及浅水域中多工况操纵运动模拟计算[J]. 中国造船, 1989(3): 38-52.

Classification method of ship navigation condition based on clustering analysis

TAN Xiao, GUANG Wenyan, LI Han, LI Yongjie, XUE Chen

Systems Engineering Research Institute, China State Shipbuilding Corporation Limited, Beijing 100036, China

Abstract With the development trend of intelligent ship and shipping as well as accumulation of ship big data, it is urgent to build a special model of navigation economic analysis through data-driven means to solve the problem of energy consumption evaluation and optimization and maximize ship energy efficiency. In this paper, combined with the relationship between ship's sailing conditions and main engine's fuel consumption characteristics, and considering the factors of draft and relative wind speed, K-means clustering analysis method is used to realize the division of different sailing conditions of the influencing factors of fuel consumption. The historical data of a VLCC are used to verify the actual application. Based on the data of ship's voyage loading and external environment weather conditions, the classification analysis of navigation conditions and the determination of influencing factor parameter interval under each working condition are realized, which provides a more refined analysis basis for the construction of matching model of marine main engine fuel consumption by different working conditions.

Keywords ship; low-speed diesel engine; fuel consumption; condition classification; clustering analysis ●



(责任编辑 王志敏)