

中文引用格式:李苗,李凌波,左志恒,等. 基于机器学习的成品油管道运行工况识别[J]. 中国安全科学学报,2024,34(6):127-135.

英文引用格式:LI Miao, LI Lingbo, ZUO Zhiheng, et al. Machine learning-based recognition for recognizing operating conditions of multi-product pipelines[J]. China Safety Science Journal, 2024, 34(6): 127-135.

# 基于机器学习的成品油管道运行工况识别\*

李苗<sup>1</sup>高级工程师, 李凌波<sup>1</sup>工程师, 左志恒<sup>1</sup>高级工程师, 张丽<sup>2</sup>,  
江路鑫<sup>3</sup>, 苏怀<sup>\*\*4</sup>副教授

(1 国家石油天然气管网集团有限公司 华南分公司, 广东 广州 510623; 2 昆仑数智科技有限责任公司 智慧天然气与管道事业部, 北京 102206; 3 国家石油天然气管网集团有限公司 科学技术研究总院分公司, 河北 廊坊 065000; 4 中国石油大学(北京) 油气管道输送安全国家工程实验室/城市油气输配技术北京市重点实验室, 北京 102249)

中图分类号: X944.4

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2024.06.1410

基金项目: 国家石油天然气管网集团有限公司科技项目(GWHT20210025353); 国家自然科学基金青年科学基金资助(51904316); 中国石油大学(北京)科学基金资助(2462021YJRC013)。

**【摘要】** 为改善成品油管道系统生产运行监测过程中不能自动识别部分运行状态, 以及异常工况识别准确率较低的问题, 应用智能工况识别方法, 构建具有实时监测能力的成品油管道运行工况识别模型。首先, 采用逻辑规则判别方法, 并根据成品油管道系统中的事件日志补充数据标签; 其次, 按照工况的起止时间对数据进行分段, 并采用滑动窗口的方式提取不同工况的子序列及其特征; 然后构建成品油管道运行工况识别模型, 并与随机森林(RF)、自适应提升(AdaBoost)、支持向量机(SVM)、时间序列森林(TSF)、随机区间谱系森林(RISF)和序列学习器(SEQL)等6种机器学习分类模型进行对比, 分析其识别效果; 最后, 以某真实成品油管道为例, 进行模型验证。结果表明: TSF模型对阀门开关、阀门内漏、清管和甩泵4种工况的识别精确度最高, 且更适合短期内运行工况的识别; 而AdaBoost模型的识别精确度在95%的置信区间内所含真实值的概率更高。

**【关键词】** 机器学习; 成品油管道; 运行工况; 工况识别; 分类模型

## Machine learning-based recognition for recognizing operating conditions of multi-product pipelines

LI Miao<sup>1</sup>, LI Lingbo<sup>1</sup>, ZUO Zhiheng<sup>1</sup>, ZHANG Li<sup>2</sup>, JIANG Luxin<sup>3</sup>, SU Huai<sup>4</sup>

(1 South China Company, China Oil & Gas Pipeline Network Corporation, Guangzhou Guangdong 510623, China; 2 Smart Gas & Pipeline Division, Kunlun Digital Technology Co., Ltd., Beijing 102206, China; 3 PipeChina Science and Technology Institute, China Oil & Gas Pipeline Network Corporation, Langfang Hebei 065000, China; 4 National Engineering Laboratory for Pipeline Safety, Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum, Beijing 102249, China)

**Abstract:** In order to solve the problems that some operating conditions could not be automatically

\* 文章编号: 1003-3033(2024)06-0127-09; 收稿日期: 2023-12-20; 修稿日期: 2024-03-21

\*\* 通信作者: 苏怀(1991—), 男, 北京人, 博士, 副教授, 主要从事复杂油气管网可靠性评价及其智能化技术研究。E-mail: suhuai1990@163.com。

identified and the accuracy of abnormal operating condition recognition was low in the process of monitoring the production and operation of multi-product pipeline system, the intelligent operating condition recognition method was applied to construct a multi-product pipeline operating condition recognition model with real-time monitoring capability. First, logic rule discrimination methods and event logs in the multi-product pipeline system were used to supplement the data labels. Second, the data were segmented according to the start and end time of the operating conditions, and the subsequence of different operating conditions were extracted by using the sliding window. Third, the features of subsequence were extracted to construct the model for operating condition recognition of multi-product pipelines, and the recognition effects of six classification models, namely, random forest (RF), adaptive boosting (AdaBoost), support vector machine (SVM), time series forest (TSF), random interval spectral forest (RISF) and sequence learner (SEQL), were compared and analyzed. Finally, a real multi-product pipeline was used as an example for model validation. The results show that the TSF model has the highest recognition accuracy for the four operating conditions of valve switching, valve internal leakage, pigging and sling pump, and is more suitable for the recognition of short-term operating conditions. In contrast, the recognition precision of the AdaBoost model has a higher probability of including the true value in the 95% confidence interval.

**Keywords:** machine learning; multi-product pipeline; operating conditions; operating condition recognition; classification model

## 0 引言

国家发改委《中长期油气管网规划》<sup>[1]</sup>中明确指出,到2025年,我国成品油管网总长将达到4万km,实现全国100万人口以上城市成品油管道的接入。对于长距离复杂管网,调度操作十分频繁,设备在运行过程中难免会出现故障,完全依靠人工经验判断工况并操作,容易出现判断不准或操作有误等情况<sup>[2]</sup>。因此,发展成品油管网工控系统大数据分析关键技术,建立基于管网关键运行参数的异常工况感知模型,可大幅度减少人为失误,提升成品油管网运行与管理水平。

目前,管道运行工况识别的研究主要聚焦于管道泄漏工况,常用的管道泄漏识别方法侧重压力、流量等实时运行数据,通过负压波法<sup>[3]</sup>或数据模型定位泄漏位置,且已经形成适用于不同管道系统的商业检漏软件;但由于导致压力变化的原因较多,管道的泄漏种类也较多,所以,现有商业检漏软件存在响应时间长、误报率和漏报率较高的弊端<sup>[4]</sup>。因此,众多学者采用机器学习算法改进现有方法,智能识别管道泄漏工况。在纯数据驱动的泄漏检测方面,戚元华等<sup>[5]</sup>通过提取正常工况下管道压力信号的统计特征,建立了基于支持向量数据描述的泄漏检测模型,用以检测天然气管道的异常信号。ZUO Zhonglin等<sup>[6]</sup>构建了基于长短期记忆网络与自编码器的混合模型,将提取到的数据特征输入支持向量

机(Support Vector Machines, SVM)模型进行训练,从而识别管道泄漏工况。文献[5-6]不需要考虑异常数据的样本量,但由于没有或较少利用异常标签信息来学习表征,可能存在着偏离真实特征的风险。为改善这个弊端,ZADKARAMI等<sup>[7]</sup>通过模拟管道泄漏工况,获取了不同泄漏条件下的压力和流量数据,提出了基于统计学和小波变换的特征提取方法,并将统计特征和小波特征作为多层感知器的输入,进而识别管道泄漏情况。ZANG Dong等<sup>[8]</sup>将时间序列转化为马尔可夫链形式,将马尔可夫特征和统计特征共同作为特征向量,这样可提高管道泄漏数据的检测准确度。在基于声发射法的泄漏检测方面,NING Fangli等<sup>[9]</sup>基于集合经验模式分解和相关系数分析声学特征,将提取到的天然气管道泄漏工况时域特征和波形特征组合后,输入到随机森林(Random Forest, RF)分类器中进行训练,即可提高管道泄漏工况的分类检测精度。赵利强等<sup>[10]</sup>采用二代小波插值法改进了经验模态分解过程,并利用其定位泄漏信号的极值点,有效去除错误的固有模态函数,提高了对管道泄漏特征的提取精度,进而提升了管道泄漏工况的识别准确率。郎宪明等<sup>[11]</sup>构建了基于双SVM的分类识别模型,对于存在多个泄漏点的管道泄漏工况,该模型的识别精度可达97%。

综上所述,现有管道泄漏工况识别方法初步解决了对某一时刻系统特定故障的辨识,但检测到的异常工况单一,对不同运行工况的泛化能力不足,鲜

有除管道泄漏工况以外的其他非稳态运行工况识别的报道。鉴于此,笔者拟综合考虑智能识别方法的现场可实施性和工况识别准确度的可验证性,提出成品油管道运行工况识别方法,探究不同特征提取方式和机器学习分类模型(RF、自适应提升(Adaptive Boosting, AdaBoost)、SVM、时间序列森林(Time Series Forest, TSF)、随机区间谱系森林(Random Interval Spectral Forest, RISF)和序列学习

器(Sequence Learner, SEQL)对工况分类准确性的影响,以期为成品油管道系统的数字化转型建设奠定基础。

### 1 成品油管道运行工况识别方法

成品油管道系统运行工况识别主要包括数据采集与预处理、数据分段、特征提取和运行工况识别等 4 部分,具体实施方法流程如图 1 所示。

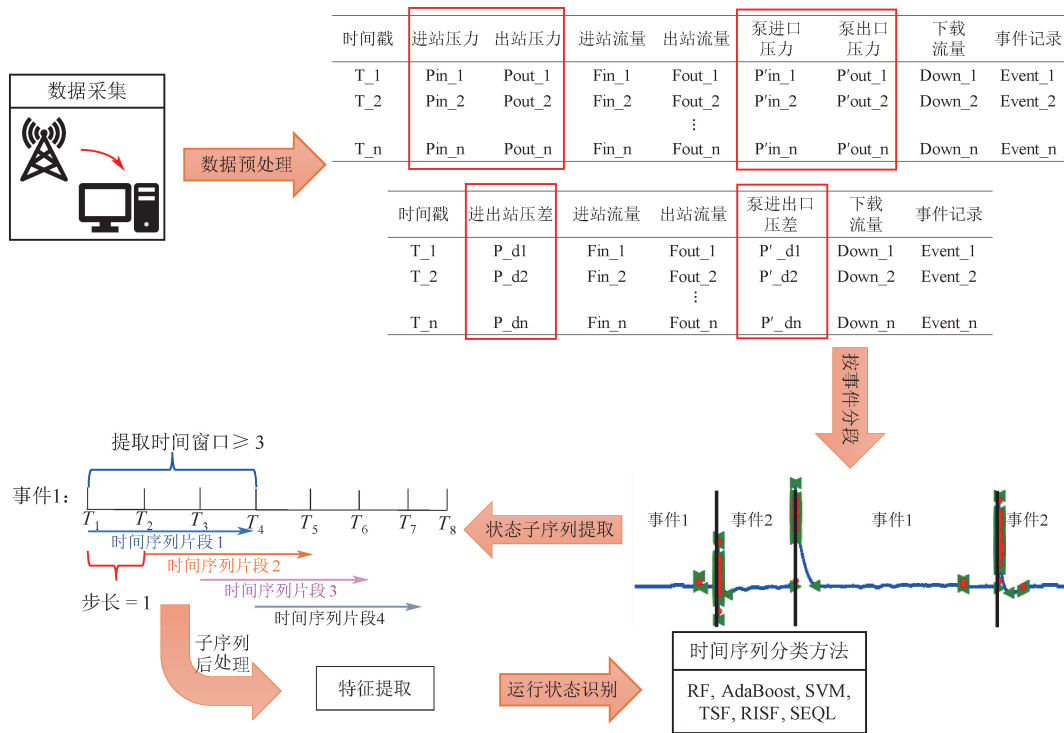


图 1 识别方法实施流程

Fig. 1 Architecture of proposed method

#### 1.1 数据预处理与分段方法

由于从监控与数据采集(Supervisory Control and Data Acquisition, SCADA)系统中采集到的数据是倒序的,首先应采用重组时间戳的方法,将数据按正序排列,当遇到重复的时间戳时,保留第 1 次出现的时间戳;然后,对于传感器传输故障导致的数据空缺问题,文中采用向前填充的方法填补缺失的数值,从而最大限度地减少原始数据的损失;最后,由于 SCADA 系统信息传输的异步性,历史事件中会出现同一时间点对应不同的运行事件的记录。基于此,文中使用正则表达式进行文本匹配,找到与案例所需识别运行工况相关的事件描述,消除其他记录的影响<sup>[12]</sup>。

将相关数据按照事件的起始时间进行分段,得到不同事件下的目标时间序列。然后,以一定时间长度提取目标时间序列中的状态子序列。文中以

60 s 的时间间隔为基准,采用滑动窗口在目标时间序列中选取所有子序列,以避免错过任何运行状态,时间步长取 1 s。

#### 1.2 分类方法与评价指标

机器学习算法的原理是通过运用大量历史数据训练模型,挖掘出其中隐含的规律,用于未来数据的预测或分类。文中用的算法均为监督学习算法,不仅要输入数据特征,还需给出数据标签,通过训练数据特征,与标签之间产生联系,从而达到预测或分类的目的。

文中主要采用 2 类机器学习模型,一类为经典机器学习模型,包括 RF、AdaBoost 以及 SVM3 种算法,此类模型主要通过提取统计特征进行分析预测;另一类为时间序列分类模型,包括 TSF、RISF、SEQL 等 3 种算法,可专门处理各类时间序列分类问题。文中所采用的 6 种分类模型的优劣势对比见表 1。

表 1 6 种分类模型的优劣势对比

Table 1 Comparison of advantages and disadvantages of six classification models

| 分类模型                     | 优势                                                               | 局限性                                                         |
|--------------------------|------------------------------------------------------------------|-------------------------------------------------------------|
| RF <sup>[13]</sup>       | 准确率高;可很好处理缺失值;适用高维数据;可度量样本间的相似性 <sup>[13]</sup>                  | 不可解释性强;在某些噪声较大的分类问题上会过拟合;准确度依赖数据体量 <sup>[13]</sup>          |
| AdaBoost <sup>[13]</sup> | 具有很高的精度;充分考虑每个分类器的权重;不用对特征进行筛选 <sup>[13]</sup>                   | 弱分类器数目不好设定;数据不平衡将导致分类精度下降;训练耗时较长 <sup>[13]</sup>            |
| SVM <sup>[13]</sup>      | 可避免维数灾难;泛化能力强;算法简单,具有较好鲁棒性;对小样本、非线性及高维模式识别具备独特优势 <sup>[13]</sup> | 难以训练大规模数据集;对参数和核函数选择敏感;实际中需要组合多个二分类实现多分类 <sup>[13]</sup>    |
| TSF <sup>[14]</sup>      | 有可解释性,计算效率高,可提取重要时间特征;可克服间隔特征空间巨大的问题 <sup>[14]</sup>             | 适用于单变量时间序列,对于多维时间序列数据的分类问题,需要先将多维数据降维,再进行训练 <sup>[14]</sup> |
| RISF <sup>[14]</sup>     | 算法简单;可准确处理噪声较大的数据 <sup>[14]</sup>                                | 用于长序列分析时,此方法运行较为缓慢 <sup>[14]</sup>                          |
| SEQL <sup>[15]</sup>     | 有可解释性,可将数字矢量转换为符号表示;可减少时间序列长度的长度矢量;更适合处理多分类问题 <sup>[15]</sup>    | 模型较为复杂,训练时间较长 <sup>[15]</sup>                               |

文中采用滑动窗口的方式将时间序列数据切成固定长度的子序列,然后对各子序列进行特征表征。值得注意的是,文中所研究的管道运行数据与经典时间序列数据不同,不存在周期性,所以在数据表征过程中不考虑时间序列的周期性。因此,文中所采用的统计特征包括:平均值、峰度、偏度、标准偏差值、方差、复杂性、近似熵和绝对能量,这些特征能够反映子序列的振幅信息,进而使分类过程能够考虑时间序列的复杂度、波动性和趋势性等特点。

文中所用 6 种模型的输入特征都直接或间接的使用了上述统计特征。其中,RF、AdaBoost 和 SVM 模型直接利用不同子序列的统计特征进行分类。TSF 模型的输入特征除了考虑统计特征外,还可以通过随机选择少数特征子集并计算其标准差来生成引导特征,以便更好地探索时间序列数据的多样性。RISF 模型利用时间序列数据的频域特征来提高性能,例如:功率谱密度、频谱包络和频率分量的均值、方差、斜度等,进而从频域的层面提高模型的性能和表达能力。SEQL 模型是通过不同子序列的统计特征提供序列分布和形态信息,然后采用符号表示不同子序列。

为了验证模型的有效性,比较不同模型的识别效果,采用 4 个评价指标评价识别结果:准确率 (Accuracy,  $A$ )、精确率 (Precision,  $P$ )、召回率 (Recall,  $R$ ) 和  $F_1$  值。其计算方法如以下 4 式<sup>[16]</sup>:

$$A = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = 2 \frac{P \times R}{P + R} \quad (4)$$

式中:TP 为对实际正样本预测正确的样本数量;TN 为对实际负样本预测正确的样本数量;FP 为对正样本预测错误的样本数量;FN 为对负样本预测错误的样本数量。

## 2 不同工况下多种模型的对比分析

### 2.1 成品油管道系统介绍与运行数据采集

模型的准确度是衡量算法是否合适最重要的特征,现以我国南方某真实成品油管道系统为例,该成品油管道总长度为 495.1 km,由 4 个泵站,3 个分输站和 2 个末站组成,各站均设有清管流程、计量设备和不同类型的阀门若干,该成品油管道系统拓扑结构如图 2 所示。受限于现场数据和事件标签约束,仅以阀门开关、阀门内漏、清管和甩泵 4 种成品油管道设备运行工况为例进行效果验证。

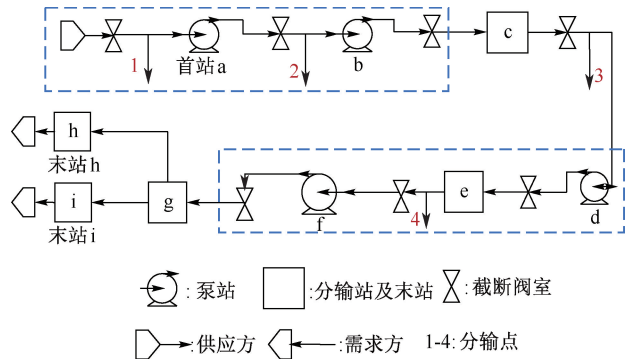


图 2 成品油管道系统拓扑结构

Fig. 2 Topological structure of pipeline system

从该成品油管道的 SCADA 系统中收集总时间长度为 2 个月 (2020 年 3—4 月和 2020 年 10—11 月) 的

历史运行数据和历史事件记录,采样频率为 1 s。收集的参数类型主要有:各站的进出口压力、泵和阀门的进出口压力等,各站所采集参数种类汇总见表 2。

表 2 各站信号信息汇总

Table 2 Summary of signal information of each station

| 站名 | 数据维度 | 采样时长/d | 采样频率/s | 信号类型                                                          |
|----|------|--------|--------|---------------------------------------------------------------|
| a  | 21   | 60     | 1      | 进出站压力、进出站流量、下载流量、主输泵(1号、2号、3号、4号)进出口压力、给油泵(1号、2号、3号、3a号)进出口压力 |
| b  | 13   |        |        | 进出站流量、进出站压力、下载流量、主输泵(1号、2号、3号、4号)进出口压力                        |
| c  | 4    |        |        | 越站流量、进出站压力、下载流量                                               |
| d  | 7    |        |        | 进出站流量、进出站压力、下载流量、主输泵进出口压力                                     |
| e  | 5    |        |        | 进出站流量、进出站压力、下载流量                                              |
| f  | 7    |        |        | 进出站流量、进出站压力、下载流量、主输泵进出口压力                                     |
| g  | 5    |        |        | 进出站流量、进出站压力、下载流量                                              |
| h  | 3    |        |        | 进站流量、进站压力、下载流量                                                |
| i  | 2    |        |        | 下载流量、进站压力                                                     |

采集到的历史事件包括各个时刻管道或设备的运行状态、操作和报警信息等。其中,清管和阀门开关动作存在现场真实记录,而阀门内漏和甩泵工况缺乏数据标签。对于无法通过 SCADA 系统直接获取的运行事件,文中提出了将人工判断经验转化为逻辑判别语句的方法。

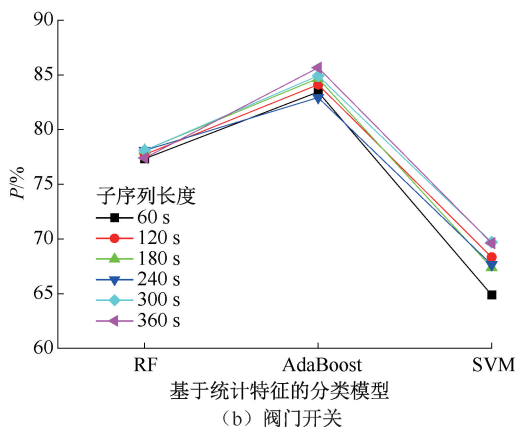
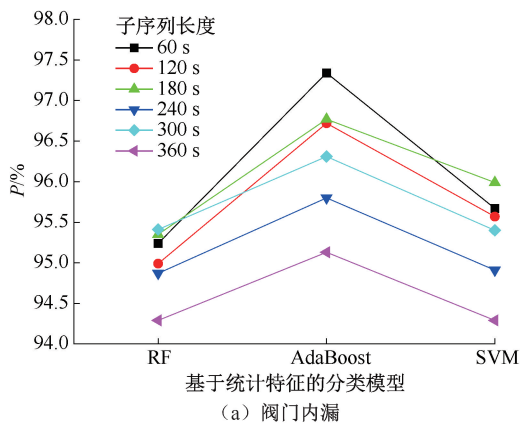
文中主要对主干线阀门内漏、甩泵和下载线阀门内漏 3 种事件进行标签补充。首先,对不同工况的人工判别经验进行梳理总结;然后,提出了基于专家经验的逻辑规则判别方法,以丰富数据信息,补充训练数据集中未标注的数据标签;最后,对不同运行工况的伪算法程序进行编制。在规则的构建程序中,主要使用了文本匹配方法和数据变化趋势对符合判别规则的时间点进行标注。

### 2.2 基于统计特征的分类模型对比试验

以提取相同统计特征作为基线,比较各类机器学习模型。基于统计特征的机器学习模型识别阀门开关动作、阀门内漏、清管和甩泵 4 种场景下运行状态,识别结果如图 3 所示。需要注意的是,对于管道运行工况的判断来说,应做到快速准确的识别,才能规避危险事故的发生。所以,当识别准确度受子序列影响较小时,模型应当选择较短的子序列长度。

从图 3 可以看出,在阀门内漏状态检测下,AdaBoost 模型的识别精度最高,且  $P$  几乎与输入子序列长度成正比。在阀门开关动作检测下,AdaBoost 模型仍然是  $P$  最高的模型。当测试子序列长度分别为 360 和 60 s 时,所得  $P$  差值仅为 0.022 1,所以此状态与输入子序列长度的相关性较

弱。在清管状态下,3 种算法  $P$  随输入子序列长度的变化而变化,当子序列长度为 360 s 时,RF 模型的  $P$  为 84.91%;当子序列长度为 60 s 时,SVM 模型的  $P$  为 84.68%,二者仅相差 0.002 3。此时,更倾向于选择测试子序列长度更小的 SVM 模型。在甩泵状态下,AdaBoost 模型得到的分类精确率最高,且几乎不受输入子序列长度的影响,此时可以选择 60 s 作为子序列输入长度。



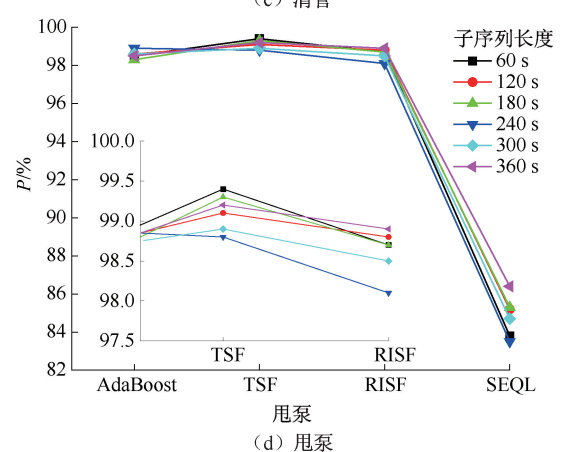
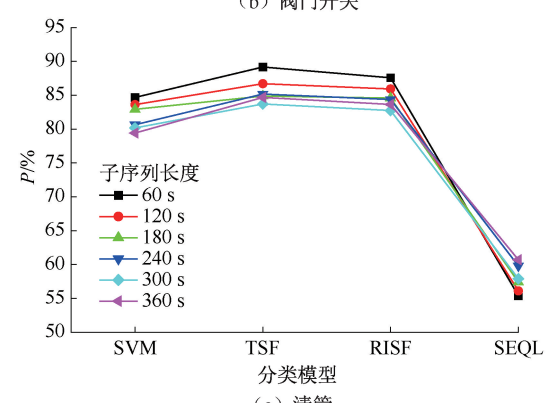
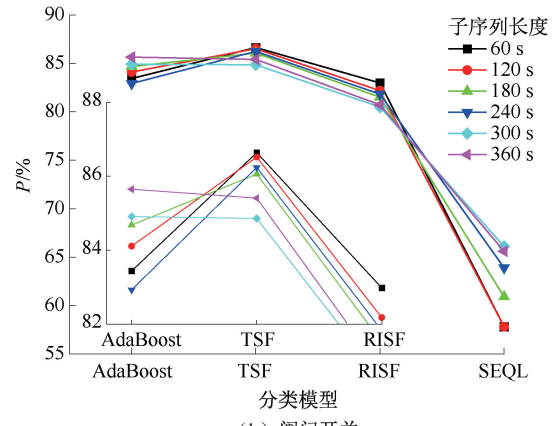
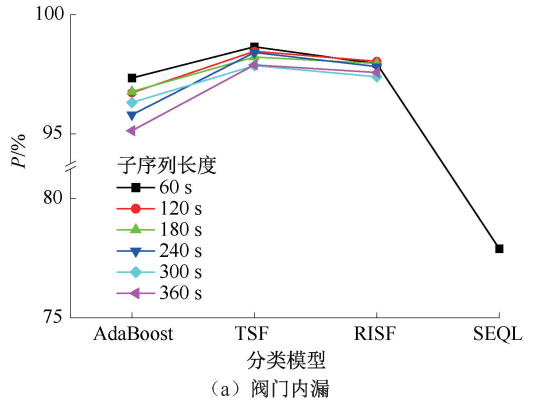
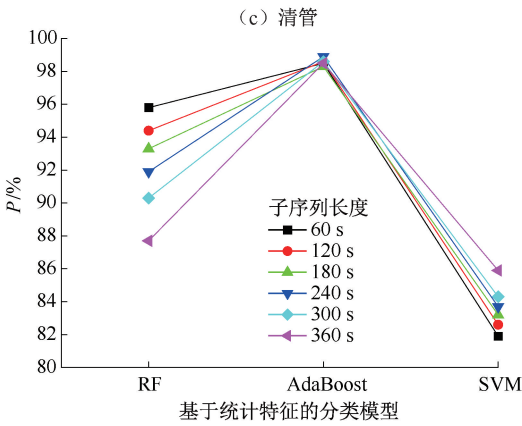
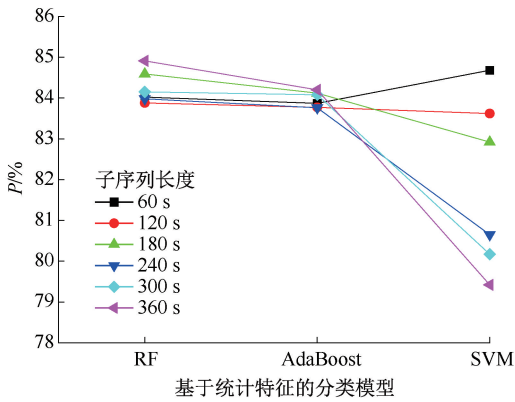


图3 RF、AdaBoost 和 SVM 模型对管道不同运行工况的识别效果

Fig. 3 Effectiveness of RF, AdaBoost and SVM models in recognizing different operating conditions of pipelines

### 2.3 基于时间序列的分类模型对比试验

以 2.2 节中得到的最优的模型作为对照组,基于时间序列特征的模型作为试验组。试验组和对照组模型对阀门内漏状态,阀门开关状态,清管和甩泵 4 种场景的识别效果对比如图 4 所示。

由图 4 可知:TSF 模型对 4 种运行工况的识别精确率是最高的。这是因为 TSF 模型的输入特征包括时间序列数据、数据窗口、统计特征、引导特征,这些特征的组合使模型能够更好地捕捉时间序列数据的模式和特征;此外,该方法引入增益余量作为分割标准选择子序列的分割点,克服了间隔特征空间巨大的问题。而其他模型只是将每个时间窗口视为一个独立的特征。

在 4 种运行工况下,SEQL 模型的识别精度最低,且识别精确率与子序列长度成正比。这是因为 SEQL 模型的核心是采用符号表示的方法,将不同时间段内的子序列转换为符号进行表示。然而,在符号表示过程中,一些较小的数据波动会被符号区

图4 不同分类模型对管道不同运行工况的识别效果  
Fig. 4 Effectiveness of different classification models in recognizing different operating conditions of pipelines

间所忽略。这就容易出现不同运行工况的符号表示结果相同的现象,进而导致该模型的识别精度相对较低。从输入子序列长度对模型识别效果的影响来看,SEQL 模型的识别精度随输入子序列长度的增大而增大。这是因为,序列越长,不同符号所代表的区间范围可能会越小,进而符号表示结果就越精准。但总的来说,该模型不适用于管道运行工况的识别。

RISE 模型主要通过提取子时间序列的光谱特征进行表征。由图 4 可知:其识别精确率低于 TSF 模型,但不易受到子序列长度的影响。这可能是因为,虽然管道运行数据在数值上存在差异,但当模型将其转换到频谱区间后,频域数据并不存在明显波动,所以此时的特征表征方法不能精准的描述某段时间内的工况变化。这也说明了基于单一频域特征的特征方法并不能很好地对成品油管道运行数据进行全面表示。

### 2.4 模型结果的可信度分析

由上述分析可知:在基于统计特征的分类模型中,阀门内漏工况的最佳识别模型是 AdaBoost 模型,最优输入子序列长度为 60 s;阀门开关工况的最佳识别模型是 AdaBoost 模型,最优输入子序列长度为 360 s;甩泵的最佳识别模型是 AdaBoost 模型,最优输入子序列长度为 240 s;清管工况的最佳识别模型是 SVM 模型,最优输入子序列长度为 60 s。在基于时间序列特征的分类模型中,4 种工况的最佳识别模型都是 TSF 模型,最优输入子序列长度均为 60 s。最优子序列长度下,2 种最佳识别方法在不同场景下的评估结果见表 3 和表 4。

表 3 不同识别方法对清管工况的评价结果

Table 3 Evaluation results of different recognition methods in pigging condition

| 应用场景 | 评价指标    | 模型对比 |      |
|------|---------|------|------|
|      |         | SVM  | TSF  |
| 清管   | $P/\%$  | 85   | 89   |
|      | $R/\%$  | 27   | 53   |
|      | $A/\%$  | 99   | 99   |
|      | $F_1$ 值 | 0.41 | 0.66 |

从表 3 可以看出,2 种模型对清管工况的识别效果差异较大,基于时间序列特征的 TSF 模型的识别精确率较 SVM 模型高 0.4,同时,召回率和  $F_1$  值均高于 SVM 模型。对于清管工况,选择 TSF 模型的漏报率更低,识别效果更好。从表 4 可以看出,AdaBoost 模型和 TSF 模型在 3 种场景下的识别精确率相差均在 0.01 左右,同时,召回率和  $F_1$  值均差别

表 4 不同识别方法在不同场景下的评价结果

Table 4 Evaluation results of different recognition methods in different scenarios

| 应用场景   | 评价指标    | 模型对比     |      |
|--------|---------|----------|------|
|        |         | AdaBoost | TSF  |
| 阀门内漏工况 | $P/\%$  | 97       | 98   |
|        | $R/\%$  | 73       | 79   |
|        | $A/\%$  | 99       | 99   |
|        | $F_1$ 值 | 0.83     | 0.87 |
| 阀门开关工况 | $P/\%$  | 98.9     | 99.3 |
|        | $R/\%$  | 75       | 77   |
|        | $A/\%$  | 99       | 99   |
|        | $F_1$ 值 | 0.85     | 0.87 |
| 甩泵     | $P/\%$  | 86       | 87   |
|        | $R/\%$  | 53       | 55   |
|        | $A/\%$  | 99       | 99   |
|        | $F_1$ 值 | 0.66     | 0.67 |

不大。所以,当出现评价结果相差不大的情况时,不能直接通过精确率或召回率来选择最优的识别模型。

为作进一步选择,采用模型识别精确率在 95% 的置信区间内所含真实值的概率进行模型识别结果可信度对比。即当出现识别精确率差异较小的情况时,可选择置信区间内所含真实值概率更高的模型进行分类。文中分别为表 4 中 3 种工况的最优识别模型建立了 95% 的置信区间。2 种模型的识别精度(在最优输入子序列长度下)在 95% 的置信区间内包含真实值的概率见表 5。

表 5 识别精确率在 95% 的置信区间内包含真实值的概率

Table 5 Probability of recognition precision containing true value within 95 % confidence interval

| 场景   | 分类模型     | 测试数据长度/s | 精确率 (95% 置信区间)/% |
|------|----------|----------|------------------|
| 阀门开关 | AdaBoost | 360      | 85.65 (±1.9)     |
|      | TSF      | 60       | 86.64 (±0.83)    |
| 甩泵   | AdaBoost | 240      | 98.9 (±0.9)      |
|      | TSF      | 60       | 99.4 (±0.2)      |
| 阀门内漏 | AdaBoost | 60       | 97.34 (±0.5)     |
|      | TSF      | 60       | 98.65 (±0.4)     |

从表 5 可以看出,基于统计特征的 AdaBoost 模型在 95% 置信区间内的可信度相对较高。具体来说,在阀门开关工况下,当  $P$  在 83.75%~87.55% 范围内时,AdaBoost 模型的  $P$  有 95% 的概率是真实的。在甩泵工况下,当  $P$  在 98%~99.8% 范围内时,AdaBoost 模型的  $P$  有 95% 的概率是真实的。在阀

门内漏工况下,当  $P$  在 96.84%~97.84% 范围内时, AdaBoost 模型的  $P$  有 95% 的概率是真实的。所以,从可信度的角度来说,AdaBoost 模型识别结果中包含真实值的概率更高,识别效果更强。

### 3 结 论

1) 文中所构建的成品油管道工况识别模型,可以准确识别管道多种非稳态运行工况,保障管道系统生产安全,助力调度人员快速决策。

2) 在基于统计特征的分类模型中,阀门内漏、阀门开关和甩泵 3 种工况的最佳识别模型均为

AdaBoost 模型,精确度分别达到 97%,98% 和 86%;清管工况的最佳识别模型是 SVM 模型,识别精度为 85%。在基于时间序列特征的分类模型中,TSF 为最佳识别模型,4 种工况的识别精度分别达到 98%、99%、87% 和 89%。

3) 当输入工况时长为 60 ~ 360 s 时,AdaBoost 模型的识别精度在 95% 置信区间内所包含真实值的概率高于 TSF 模型,因此,当模型精度相近时,应优先选择输入子序列长度灵活和可信度更高的 AdaBoost 模型。

### 参 考 文 献

- [1] 国家发展改革委,国家能源局. 中长期油气管网规划[Z]. 2017.
- [2] 张丽. 基于机器学习的成品油管道系统运行工况识别方法研究[D]. 北京:中国石油大学, 2023. 5.  
ZHANG Li. Research on machine learning based method for operational states recognition of product oil pipeline systems[D]. Beijing: China University of Petroleum, 2023. 5.
- [3] 秦程. 基于负压波与流量平衡法的管道泄漏监测系统研究[D]. 大连:大连理工大学, 2021.  
QIN Cheng. Research on pipeline leakage monitoring system based on negative pressure wave and flow balance method[D]. Dalian: Dalian University of Technology, 2021.
- [4] ABHULIMEN K E, SUSU A A. Liquid pipeline leak detection system: model development and numerical simulation[J]. Chemical Engineering Journal, 2004, 97(1): 47-67.
- [5] 戚元华,林伟国,吴海燕. 基于时域统计特征的天然气管道泄漏检测方法[J]. 石油学报, 2013, 34(6): 1 195-1 199.  
QI Yuanhua, LIN Weiguo, WU Haiyan. Natural gas pipeline leakage detection method based on time-domain statistical characteristics[J]. Acta Petrolei Sinica, 2013, 34(6): 1 195-1 199.
- [6] ZUO Zhonglin, MA Li, LIANG Shan, et al. A semi-supervised leakage detection method driven by multivariate time series for natural gas gathering pipeline[J]. Process Safety and Environmental Protection, 2022, 164: 468-478.
- [7] ZADKARAMI M, SHAHBAZIAN M, SALAHSHOOR K. Pipeline leak diagnosis based on wavelet and statistical features using Dempster-Shafer classifier fusion technique [J]. Process Safety and Environmental Protection, 2017, 105: 156-163.
- [8] ZANG Dong, LIU Jinhai, WANG Huaizhen. Markov chain-based feature extraction for anomaly detection in time series and its industrial application[C]. 2018 Chinese Control and Decision Conference (CCDC). IEEE, 2018: 1 059-1 063.
- [9] NING Fangli, CHENG Zhanghong, MENG Di, et al. A framework combining acoustic features extraction method and random forest algorithm for gas pipeline leak detection and classification[J]. Applied Acoustics, 2021, 182: DOI: 10.1016/j.apacoust.2021.108255.
- [10] 赵利强,王建林,于涛,等. 基于二代小波变换和多级假设检验的输油管道缓泄漏检测方法[J]. 石油学报, 2012, 33(5): 898-903.  
ZHAO Liqiang, WANG Jianlin, YU Tao, et al. Slow leakage detection method of oil pipeline based on second-generation wavelet transform and multi-stage hypothesis test[J]. Acta Petrolei Sinica, 2012, 33(5): 898-903.
- [11] 郎宪明,王佳政,曹江涛,等. 基于改进 VMD 和 TWSVM 的多点泄漏检测方法[J]. 振动与冲击, 2021, 40(17): 271-278.  
LANG Xianming, WANG Jiazheng, CAO Jiangtao, et al. Multi-point leak detection method based on improved VMD and TWSVM[J]. Journal of Vibration and Shock, 2021, 40(17): 271-278.
- [12] 张丽,苏怀,范霖,等. 基于时序片段的油气管道运行工况识别方法[J]. 中国安全生产科学技术, 2022, 18(11):

99-104.

ZHANG Li, SU Huai, FAN Lin, et al. Identification method of oil and gas pipeline operating conditions based on time series fragments[J]. Journal of Safety Science and Technology, 2022, 18(11): 99-104.

- [13] KHASHA R, SEPEHRI M M, TAHERKHANI N. Detecting asthma control level using feature-based time series classification[J]. Applied Soft Computing, 2021, 111: DOI: 10.1016/j.asoc.2021.107694.
- [14] LINES J, TAYLOR S, BAGNALL A. Time series classification with HIVE-COTE: the hierarchical vote collective of transformation-based ensembles[J]. ACM Transactions on Knowledge Discovery from Data, 2018, 12(5): 52. 1-52. 35.
- [15] NGUYEN T L, GSPONER S, IFRIM G. Time series classification by sequence learning in all-subsequence space[C]. IEEE International Conference on Data Engineering. ICDE, 2017: 947-958.
- [16] 张江石,李泳傲,冒香凝,等. 基于 NLP 的煤矿事故原因分类研究[J]. 中国安全科学学报, 2023, 33(6):20-26.  
ZHANG Jiangshi, LI Yongtun, MAO Xiangning, et al. Research on classification of causes of coal mine accidents based on NLP[J]. China Safety Science Journal, 2023, 33(6): 20-26.

**作者简介:** 李苗 (1990—),男,湖北仙桃人,博士,高级工程师,主要从事管道输送工艺技术、成品油管道输送工艺与新能源技术等方面的工作。E-mail: limiao@ pipechina.com.cn。



## 《科技期刊世界影响力指数(WJCI)报告》(2023)发布 《中国安全科学学报》继续入选 Q2 区



2023年11月,《科技期刊世界影响力指数(WJCI)报告》(2023版)(以下简称《WJCI报告》)正式发布,《中国安全科学学报》入选。

《WJCI报告》是中国科协系列课题《面向国际的科技期刊影响力综合评价方法研究》等的研究成果,该系列课题旨在建立新的期刊评价系统,探索面向全球的更为科学、全面、合理的期刊影响力评价方法,为世界学术评价融入更多中国观点,中国智慧,推动世界范围内科技期刊的公平评价、同质等效使用。

在《WJCI报告》中,《中国安全科学学报》入选安全科学技术、灾害及其防治学科,在统计周期内的总被引频次为4439,影响因子为1.947,WJCI为2.038,WJCI学科排名为23/67,位于Q2区。