

# 基于运行数据的锅炉故障自动检测技术研究

郭永谦

(青海省特种设备检验所,青海 西宁 810000)

**摘要:**为实时检测锅炉的运行数据,提高其连续工作性能,提出一种基于运行数据的锅炉故障自动检测方法。通过时间序列方式收集锅炉稳态数据,并归一化处理;根据处理后数据属性的均值、方差得出数据的集合与方差矩阵,提取出其数据特征;通过标准化方式得出特征数据各点之间的距离与距离矩阵,采用欧式距离与绝对距离测量融合方式检测出分布密集数据的故障数据,并利用距离度量函数,完成分布稀疏数据的故障检测,从而完成锅炉整个故障数据自动检测。实验结果表明:该方法锅炉故障检测结果精度高,故障检测消耗时间极短,具备良好的鲁棒性。

**关键词:**锅炉;数据采集;数据特征提取;故障数据;自动检测;数据点

**中图分类号:**TP274 **文献标志码:**B **文章编号:**1671-5276(2022)04-0237-04

## Research on Boiler Fault Automatic Detection Technology Based on Operation Data

GUO Yongqian

(Qinghai Special Equipment Inspection Institute, Xining 810000, China)

**Abstract:** In order to detect the boiler operation data in real time and improve the continuous working performance, an automatic detection method of boiler fault is proposed based on operation data. The steady-state data of boiler are collected by time series and normalized. According to the mean and variance of the processed data attributes, the data set and variance matrix are obtained, and the data features are extracted. The distance and distance matrix between each point of feature data are gained through standardization. The fault data of distributed dense data is detected by fusion of euclidean distance and absolute distance measurement. The fault detection of distributed sparse data is completed by using distance measurement function. The experimental results show that the proposed method has high accuracy, short fault detection time and good robustness.

**Keywords:** boiler; data collection; data feature extraction; fault data; automatic detection; data point

## 0 引言

现阶段锅炉的使用规模不断扩大<sup>[1]</sup>,锅炉系统的复杂程度也逐渐提升。尤其是锅炉设备的日益高速化、自动化,使其结构愈来愈繁琐。在科技背景下,可将传统人工现场操作方式转换成通过计算机网络智能控制,并能收集与保存海量的运行数据。但是,工作人员极难依据这些数据真正了解锅炉系统实际工作情况,且不能及时发现系统故障并进行精准地判定。为此,单纯通过人工分析方式来解决锅炉故障问题<sup>[2]</sup>已很难满足现今运行的需求。

为此,诸多学者、专家等对系统故障自动检测技术进行研究。张维等<sup>[3]</sup>针对锅炉系统故障检测正确率较低的状况,采用关联规则挖掘方式得出正常工作情况下锅炉设备状态参考数值,并根据测试数据和参考数值间的滑动平均相似程度来断定锅炉是否发生故障。此方法有效提升了故障自动检测精度,但计算量大,需要消耗大量时间。许裕粟等<sup>[4]</sup>面对锅炉故障不能及时检测的问题,采用K-均值算法将锅炉工作流程的工况细分出来,利用距离度量完成每个工况内各采样点的协方差距离,并根据高斯阈值建立判决函数,检测出锅炉故障。该方法可以及时检测出故

障,使用计算方法较为简单,但故障查全率较低。

基于此,本文结合文献方法优点并利用运行数据方式对锅炉故障自动检测技术进行研究。通过滑动判断方式筛选出锅炉数据,使选取的数据具有全面性;利用主成分分析方法提取出数据特征,能够在保证数据提取精度前提下降低数据提取的计算量;运用距离度量方式检测出锅炉故障,可以使自动检测结果更加贴近实际情况,并且计算简便,缩短了故障检测时间,进而完成锅炉故障自动检测。

## 1 运行数据下锅炉故障自动检测

### 1.1 故障自动检测流程

基于运行数据的锅炉故障自动检测<sup>[5]</sup>流程主要分为两个部分。第一部分包括锅炉数据的采集、归一化处理与特征提取。数据采集是经过数据清洗来提升数据质量;归一化处理为了降低数据间因数据的量级、分布情况对数据特征提取所造成的影响;数据特征提取为了更加有效表述锅炉的特征与属性;第二部分是利用距离度量方式来判定故障数据,使故障检测结果更加精准,达到自动检测目的,具体过程如图1所示。

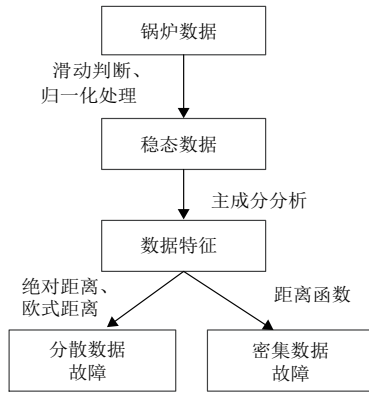


图1 故障检测流程

## 1.2 运行数据特征提取

### 1) 数据筛选

锅炉设备的工作状态是根据电网负荷的需求而变化的,其工作流程为:电网负荷调度中心把负荷指标下发到电厂,电厂依据负荷调节运行的参数,使锅炉设备达到电网负荷标准之后稳定运行;当电网负荷发生改变时,锅炉设备须进行调整。

锅炉设备运行参数主要分为两种,即根据电网负荷变化而发生改变的参数与被调控在特定区域内的参数。这两种参数影响锅炉当前的工作状态。

当锅炉设备处于过渡阶段时,锅炉内部的流场会带来极强波动,致使运行监测设备的数值剧烈变化。与此同时,因当今锅炉系统规模逐渐扩大,对应的延迟也会增加,使得不同设备测量得出时间节点数据各不相同,为此不能正确反映出锅炉的工作情况。当锅炉设备在稳定的情况下工作,锅炉内部变换较为平稳,监测设备得出数值具有良好的稳定性和同步性,并能正确展现出锅炉设备当前工作情况。

利用滑动判别方式完成稳态数据采集。滑动判别方式是根据时间序列<sup>[6]</sup>找出稳态数据,通过逐点遍历的方式,随着时间维度滑动<sup>[7]</sup>,判断出数据点是否属于稳态数据。设定  $x$  表示数据点,则数据点一维空间向量  $X$  为

$$X = [x_1, x_2, \dots, x_n] \quad (1)$$

同时,原始权值向量  $W$  为

$$W = [w_1, w_2, \dots, w_n] = [0, 0, \dots, 0] \quad (2)$$

其中  $[w_1, w_2, \dots, w_n]$  每一项均对应相应的向量  $X$ ,用于判断向量  $X$  中各元素是否属于稳态数据<sup>[8]</sup>。针对向量  $X$  中元素  $x_i$ ,则有

$$\frac{x_{i-H} + x_{i-(H-1)} + \dots + x_i}{x_i + x_{i+1} + \dots + x_{i+H}} < K \quad (3)$$

或者

$$\frac{x_{i-H} + x_{i-(H-1)} + \dots + x_i}{x_i + x_{i+1} + \dots + x_{i+H}} > \frac{1}{K} \quad (4)$$

则认定  $x_i$  是非稳态部分数据点,相反的点,认定为稳态部分数据点。如图2所示,图中  $b$  表示稳态部分数据点,  $a$  不符合式(3),  $c$  不符合式(4)则认定是非稳态数据点。

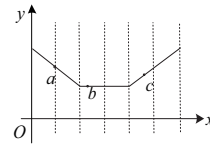


图2 滑动判定流程

若向量  $X$  中元素  $x_i$  被判定为非稳态点时,数据点  $x_{i-H} \sim i_{i+H}$  对应的权值向量  $W$  元素各加1,得出

$$[W_{i-H}, \dots, W_i, \dots, W_{i+H}] = [W_{i-H}+1, \dots, W_i, \dots, W_{i+H}+1] \quad (5)$$

式中  $i_{i+H}$  表示与  $x$  对应的非稳态点。

当向量  $X$  全部向量元素通过稳态判定后,并对权值向量  $W$  中元素进行判定,若其小于阈值,则其对应的向量  $X$  中元素  $x_i$  是稳态点,相反是非稳态点,设定为  $i_i$ 。这种方式能够通过数据点  $x_i$  的数据情况来判定数据是否属于稳态数据,使收集的数据避免了局部性错误的问题。

### 2) 数据处理

采用归一化方式将收集数据进行处理,其处理方式主要分为  $\min$ - $\max$  归一化与  $Z$ -score 归一化两种方式。数据向量  $X$  通过  $\min$ - $\max$  归一化后,得出数据向量  $x^*$

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

式中:  $x_{\min}$ 、 $x_{\max}$  表示向量  $X$  的两个极值;  $x^*$  表示进行归一化后的数据。归一化后可以把数据集合的中心移到坐标原点。

$Z$ -score 归一化处理,得出

$$x^* = \frac{x - E[x]}{\sqrt{v_{ar}[x]}} \quad (7)$$

式中:  $E[x]$  表示向量  $X$  的平均值,是均值向量;  $v_{ar}[x]$  表示向量  $X$  的方差。 $Z$ -score 归一化把数据集合的平均值变成0,方差变成1。

## 1.3 特征提取

将锅炉设备稳态处理后的数据当作数据样本集合,设定样本集合  $X \in R^{m \times n}$ ,  $m$  表示样本数量,  $n$  表示数据属性数量(工作运行参数数量)。

$$X = [X_1, X_2, \dots, X_m]^T \quad (8)$$

$$X_i = [x_{i1}, x_{i2}, \dots, x_{in}], i = 1, 2, \dots, m$$

为了解决集合中不相同属性取值结果对数据特征提取的影响,把全部属性归一化处理,使数据集合元素属性的平均值、方差分别为0、1,得出

$$\bar{X} = [X - I_n \mu] D_e^{-1/2} \quad (9)$$

式中:  $I_n = [1, 1, \dots, 1]^T$  表示向量列数为1的  $n$  个元素;  $\mu = [\mu_1, \mu_2, \dots, \mu_n]$  表示列数为  $j$  的数据平均值,也就是属性的平均值,  $j = 1, 2, \dots, n$ ;  $D_e = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$  表示对角线上各数据属性方差矩阵的集合。

通过公式(9)得出数据的集合与方差矩阵,找出协方差矩阵的  $n$  个特征数值与特征向量,即  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 、 $[p_1, p_2, \dots, p_n]$ 。关于数据集合  $\bar{X}$ ,有以下关联关系:

$$\bar{X} = TP^T = [\bar{T}, \bar{T}][\bar{P}, \bar{P}]^T = \bar{T} \bar{P}^T + \bar{T} \bar{P}^T = \bar{T} \bar{P}^T + E \quad (10)$$

式中:  $\mathbf{P} = [\bar{\mathbf{P}}, \tilde{\mathbf{P}}]$  表示特征向量矩阵;  $\bar{\mathbf{P}} = [p_1, p_2, \dots, p_a] \in R^{n \times a}$  表示主成分矩阵<sup>[9]</sup>;  $\mathbf{T}$  表示得分矩阵;  $\mathbf{E}$  表示各样本在主成分矩阵的权值。由于特征矩阵符合单位正交矩阵要求  $\mathbf{P}\mathbf{P}^{-1} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$ , 为此对于式(10)有  $\bar{\mathbf{X}}\mathbf{P} = \mathbf{T}\mathbf{P}^T\mathbf{P} = \mathbf{T}$ , 进而得出对应的得分矩阵。

一般情况下, 在主成分矩阵中特征向量数量根据特征向量的解释率与累积解释率所决定的, 取前  $a$  个解释率数值大的特征向量组成主成分矩阵。数据特征向量的解释率运算过程如式(11)所示。

$$E_R(i) = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \quad (11)$$

式中  $E_R(i)$  表示第  $i$  个数据特征向量的解释率数值。

根据特征数值大小由高到低分别算出各特征向量的解释率, 累计解释率计算过程为

$$C_{ER}(i) = \sum_{j=1}^i E_R(j) \quad (12)$$

式中:  $C_{ER}(i)$  描述了前  $i$  个特征向量的累计解释率数值;  $E_R(j)$  描述了前  $i$  个特征向量的数据变化解释占总数据变化解释的百分比。

综上所述, 归一化数据被分成主元子空间  $\bar{\mathbf{P}}$  与残差子空间  $\tilde{\mathbf{P}}$ 。通常情况下, 因为主元子空间特征向量使用较少的属性就可以将数据集合中数据大部分变换解释, 所以主成分分析能够有效提取出数据特征。与此同时, 使用较少的属性就能将数据变化的特征点变得更加突出, 也就是数据在主元子空间中坐标的方差更大, 有利于下面故障数据的检测。

## 1.4 故障自动检测

采用全局距离方式完成锅炉故障自动检测。先算出锅炉机组数据特征提取数据集合中数据两个点的间距, 与剩余点的间距, 并算出各个点到剩余点的距离和。设定  $n'$  表示故障数据数量, 则距离和最大的前  $n'$  个数据点认为是锅炉故障数据详细描述为:

将锅炉特征提取数据做标准化处理, 算出  $m$  个数据点间的距离为  $d_{ij}$ , 得出距离矩阵  $\mathbf{R}$ :

$$\mathbf{R} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & d_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mm} \end{bmatrix} \quad (13)$$

设定  $p_i = \sum_{j=1}^m d_{ij}$ ,  $p_i$  表示距离矩阵  $\mathbf{R}$  中第  $i$  行中  $d_{m1}$ ,  $d_{m2}$ ,  $\dots$ ,  $d_{mm}$  距离和, 若  $p_i$  数值越小, 表明数据点  $i$  与剩余数据点的距离和越近,  $p_i$  数值最小的前  $n'$  个数据则认定为锅炉正常数据, 反之, 为锅炉故障数值。

距离度量方式基本上采用绝对距离与欧式距离两种方式。关于  $m$  维空间中数据点  $x_i$  与  $x_j$  数点据间的欧式距离描述为

$$d_{ij}^o = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (14)$$

绝对距离描述为

$$d_{ij}^m = \sum_{k=1}^m |x_{ik} - x_{jk}| \quad (15)$$

将以上两种方式进行融合, 得出

$$d_{ij}^N = \left[ \sum_{k=1}^m |x_{ik} - x_{jk}|^q \right]^{\frac{1}{q}} \quad (16)$$

一般情况下, 数据分布为密集分布与稀疏分布, 其中呈密集分布的正常数值较多, 使周围故障数值与对应数据点间距较大。由于故障数据容易受分布密度干扰, 因此造成相反情况发生, 即稀疏周围故障数值与对应数据点间距较大。根据这一属性特征, 式(16)只能完成锅炉稀疏数据的故障自动检测, 而密集部分数据采用距离度量函数方式完成锅炉故障自动检测, 即

$$d_{ij}^U = \frac{d_{ij}^o}{\sqrt{d_{i(m)}^o \cdot d_{j(m)}^o}} \quad (17)$$

式中  $d_{i(m)}^o$ 、 $d_{j(m)}^o$  分别描述了  $x_i$ 、 $x_j$  到剩余数据的距离均值。

若数据  $s$  到数据  $u$  间的距离与数据  $t$  到数据  $w$  间的距离等同, 即

$$\begin{cases} d_{su}^o = d_{tw}^o \\ d_{s(m)}^o < d_{t(m)}^o \\ d_{u(m)}^o < d_{w(m)}^o \end{cases} \quad (18)$$

根据式(18)可知,  $d_{su}^U > d_{tw}^U$ 。

通过距离度量函数方式能够将分布密集数据的故障数据和其他数据之间的距离增大, 还能减少正常数据与其他数据之间的距离和, 进而校正故障数据与正常数据到剩余数据的距离和。为此, 可运用距离度量函数方式完成锅炉数据的故障自动检测。

## 2 实验结果分析

### 2.1 实验环境

实验整个流程在 Microsoft Windows10 操作系统下进行的, 实验平台软件选择 MATLAB R2018a 软件。

实验的数据来源于某大规模火电厂 4 号锅炉工作现场。因为电站锅炉的真实检测数据集合中可能含有故障数据, 若直接将其当作实验数据, 对得出结果不能保证其评估结果的精度。因此, 结合电站锅炉检测数据的现实收集情况, 先从锅炉机组负荷为 300 MW 的数据集合中选取 690 个数据当作正常数值, 再从锅炉机组负荷为 270 MW 中抽取 10 个数据当作故障数据, 共同组成实验 700 个数据。

### 2.2 精度对比

为了证实本文提出基于运行数据的锅炉故障自动检测技术的有效性, 从检测出的故障总数量与正确检测故障数量两个角度进行验证, 并使用对比方式将本文方法与文献[3]方法、文献[4]方法进行对比, 得出结果如图 3 所示。

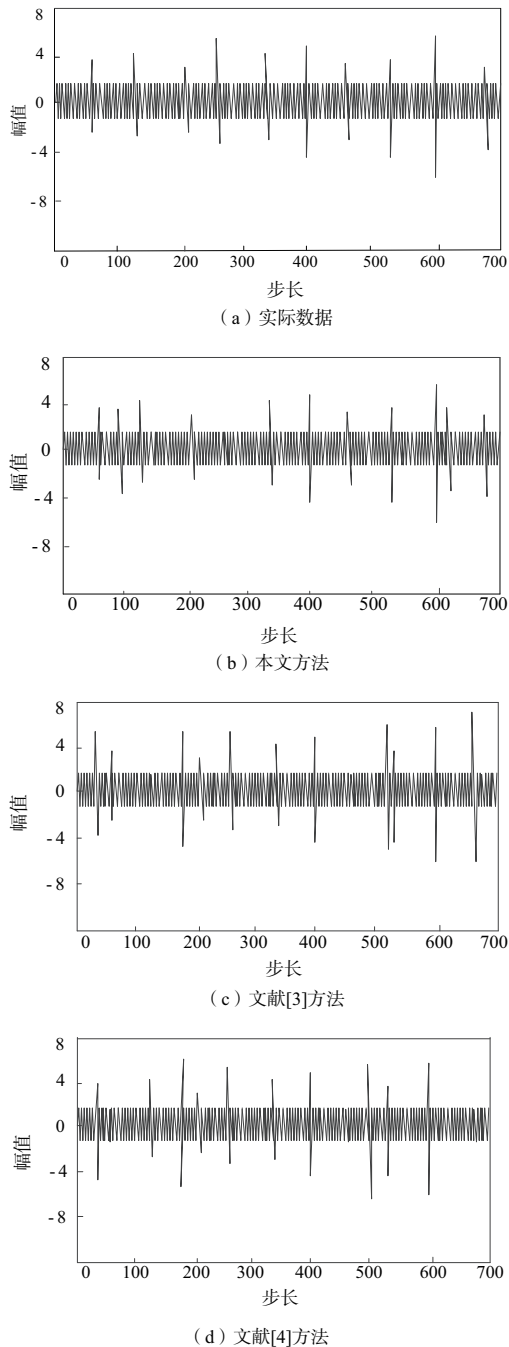


图3 检测结果对比

设定幅值在 $[2, -2]$ 之间当作是锅炉正常数据,反之当作是锅炉故障数据。图3(a)为实际锅炉运行数据,包含10个故障数据;图3(b)为本文方法锅炉故障检测结果,检测出9个正确故障数据与2个错误故障数据;文献[3]、文献[4]方法均检测出7个正常故障数据与3个错误故障数据。本文方法检测出正确、故障数据均大于文献方法,而检测出错误故障数量小于文献方法。本文方法的检测效果更好,这是因为本文方法通过时间序列方式收集锅炉稳态数据,并归一化处理;利用了主成分方式将锅炉数据进行特征提取,并运用距离度量算法完成各数据点的故障自动检测,有效缩小了故障检测结果的误差,故本文方法优于文献方法。

## 2.3 运行耗时对比

为了进一步验证3种方法的故障检测结果性能的优劣,下面从故障检测耗时角度进行验证,得出结果如图4所示。

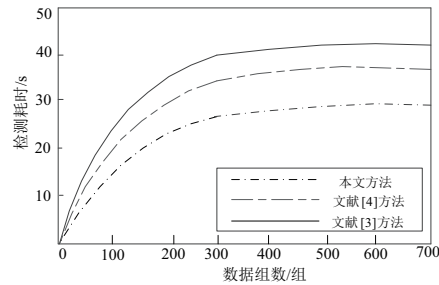


图4 故障检测耗时对比情况

从图4可以看出,当数据组数为300时,3种方法锅炉故障自动检测结果趋于平缓。数据组数为500组时,文献[3]方法、文献[4]方法与本文方法的检测耗时分别为42 s、37 s、28 s,本文检测耗时曲线均低于文献方法,说明本文方法的故障检测效率较高。这是因为由于文献方法没有使用归一化方式将锅炉数据进行处理,增加了锅炉数据特征难度,故文献方法需要消耗较多时间才能检测出锅炉故障,进一步证实了本文方法锅炉故障自动检测性能良好。

## 3 结语

锅炉实际运行数据中可能含有一些故障数据,导致锅炉设备瘫痪,不能正常运行,为此本文在运行数据基础上对锅炉故障自动检测技术进行研究。通过滑动判断方式完成锅炉数据采集,利用主成分分析方式提取出数据特征,并根据数据各点间距离判定出故障数据,完成锅炉故障自动检测。本文方法虽取得一定成效,但由于研究时间有限,日后可以将时间序列、深度学习等方法引入,进一步提升检测结果的精度。

### 参考文献:

- [1] 牟彦春,戈浩. 基于远场涡流和视频技术的电站锅炉水冷壁自动爬壁检测[J]. 锅炉技术,2020,51(5):56-60.
- [2] 侯松. 高变废热锅炉封头泄漏处理及叶式密封更换[J]. 石油化工设备技术,2020,41(5):63-66,9.
- [3] 张维,刘吉臻,高明明. 基于数据挖掘的循环流化床锅炉辅机故障预警[J]. 动力工程学报,2019,39(10):826-833.
- [4] 许裕粟,张静,李柠,等. 基于数据挖掘的锅炉在线运行状态监测[J]. 热能动力工程,2019,34(2):82-87,115.
- [5] 李亚娟,卢剑,许臻,等. 某火电厂锅炉补给水处理系统运行故障分析及解决措施[J]. 水处理技术,2020,46(2):133-136.
- [6] 崔尚勇. 锅炉燃烧的远程监控系统设计研究[J]. 自动化技术与应用,2020,39(5):63-66.
- [7] 苗壮. 基于工业锅炉的远程物联网监测数据采集终端分析[J]. 工业加热,2019,48(4):56-58,64.
- [8] 张维,刘吉臻,高明明. 基于数据挖掘的循环流化床锅炉辅机故障预警[J]. 动力工程学报,2019,39(10):826-833.
- [9] 陈江明,郑益民,段辉,等. 某型涡轴发动机鼓风机漏油故障分析[J]. 机械制造与自动化,2020,49(6):76-78.

收稿日期:2021-06-25