

基于改进 KNN 算法的汽轮机通流故障 诊断方法及应用

闫城, 陈时熠, 华心果, 向文国

(东南大学能源热转换及过程测控教育部重点实验室, 江苏 南京 210096)

[摘要] 汽轮机的故障诊断对整个电厂的安全运行意义重大。根据热力参数建立计算模型可以及早地观测到性能退化趋势, 预测设备故障类型。本文采用特征通流面积的方法建立汽轮机系统性能退化模型, 模拟系统故障样本与测试样本, 建立设备故障样本库。通过使用改进的 KNN (K -nearest neighbor) 算法, 基于汽水系统热力参数变化规律, 计算当前机组运行数据样本相对于设备故障样本的相似度, 判定当前机组各设备已发生故障的概率。通过对某 S109FA 联合循环机组汽轮机研究结果表明, 特征通流面积在不同工况下的计算误差均在 5% 以内, 满足工程计算要求。相比于传统 KNN 算法, 改进 KNN 算法通过样本评估邻居在决策过程中的权重, 取得了比传统 KNN 算法更高的分类正确率。对测试样本故障诊断结果表明, 改进 KNN 算法比传统 KNN 算法诊断准确率更高, 对测试样本诊断准确率为 100%, 采用改进 KNN 算法汽轮机系统故障诊断具有可行性, 与现场实际情况吻合。

[关键词] 联合循环; 汽轮机; 特征通流面积; KNN 算法; 故障诊断; 热力参数

[中图分类号] TK268 **[文献标识码]** A **[DOI 编号]** 10.19666/j.rlfid.202012289

[引用本文格式] 闫城, 陈时熠, 华心果, 等. 基于改进 KNN 算法的汽轮机通流故障诊断方法及应用[J]. 热力发电, 2021, 50(7): 84-90. LYU Cheng, CHEN Shiyi, HUA Xinguo, et al. Fault diagnosis method of turbine flow passage based on improved KNN algorithm and its application[J]. Thermal Power Generation, 2021, 50(7): 84-90.

Fault diagnosis method of turbine flow passage based on improved KNN algorithm and its application

LYU Cheng, CHEN Shiyi, HUA Xinguo, XIANG Wenguo

(Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: The fault diagnosis of steam turbine is of great significance to the operation of power plant. Models based on thermal parameters can be used to observe performance degradation of equipment and to predict its failure type. This paper uses the characteristic flow area method to establish the performance degradation model of the steam turbine system, simulates the system fault samples and test samples, and establishes the equipment fault sample library. Through the improved K -nearest neighbor (KNN) algorithm, based on the change law of thermal parameters of the steam and water system, the similarity of the current unit operating data samples to the equipment failure samples is calculated, and the probability of the current unit equipment failures is determined. The research shows that, the calculation error of the characteristic flow area under different operation conditions are all within 5%, which meets the calculation requirements. Compared with the conventional KNN algorithm, the improved KNN algorithm evaluates the weight of neighbors in the decision-making process through samples, and achieves a higher classification accuracy rate. The fault diagnosis results of the test sample show that, the improved KNN algorithm has a higher diagnostic accuracy rate than the conventional KNN, and the diagnostic accuracy of the test samples is 100%. The use of the improved KNN algorithm for steam turbine system fault diagnosis is feasible and consistent with the actual situation.

Key words: combined cycle, steam turbine, characteristic flow area, KNN algorithm, fault diagnosis, thermodynamic parameters

修回日期: 2021-02-03

基金项目: 国家科技重大专项(2017-I-0002-0002)

Supported by: National Science and Technology Major Project (2017-I-0002-0002)

第一作者简介: 闫城(1995), 男, 硕士研究生, 主要研究方向为联合循环机组性能分析, chenglvlc@163.com.

通信作者简介: 向文国(1964), 男, 教授, 博士生导师, wgxiang@seu.edu.cn.

众所周知,能源是人类生活中必不可少的一部分^[1]。传统的火力发电由于污染重、效率低等缺点开始逐渐被其他发电方式所替代。燃气轮机联合循环发电由于其高效率、低排放、分布式等特性,已经逐渐成为能源领域重要的发展和研究方向之一^[2]。联合循环用汽轮机是燃气-蒸汽联合循环装置中重要的设备之一,其技术水平也直接影响到联合循环机组的循环效率、设备可用率及单位造价^[3]。因此,联合循环汽轮机的经济性能、运行灵活性和安全可靠性能将影响联合循环的总体性能。

作为发电的主要设备,汽轮机长期处于高温、高压、高速的工作环境中,极易发生性能退化^[4],一旦发生退化对汽轮机的安全和经济运行将产生巨大的影响,因而对汽轮机退化诊断具有重要意义^[5]。目前美国通用电气(GE)、日本东芝电气、日立电气、富士和三菱等企业均已拥有完善的汽轮机故障诊断技术和体系。

我国在汽轮机故障诊断研究方面起步较晚,与西方发达国家有一定的差距^[6]。董晓峰基于RCM分析,采用主元分析提取通流部分故障征兆,再进行聚类分析,归纳了通流部分的故障模式类^[7-8];忻建华等根据通流故障与热力参数的关系,提出了高压缸通流部分故障的热参数模糊诊断法^[9];上海交通大学郝志莉、叶春总结了汽轮机通流部分故障的特性并进行了分类^[10]。

目前的故障诊断方法一般都只能发现汽轮机的性能发生变化,但并不能准确地判断出故障的具体位置及类型。

因此本文提出了基于特征通流面积建立汽轮机系统性能退化模型,模拟汽轮机系统的故障样本与测试样本,建立设备故障样本库。基于汽轮机系统热力参数变化规律,计算当前机组运行数据样本相对于设备故障样本的相似度,判定当前机组各设备已发生故障的概率,实现对汽轮机系统设备故障的快速诊断。

1 汽轮机系统

本文以某S109FA联合循环机组汽轮机为研究对象,该机组由1台F级燃气轮机、1台余热锅炉,1台蒸汽汽轮机组成。汽轮机为三压、再热、抽汽凝汽式、双缸双排式汽轮机,部分设计参数见表1。

1.1 通流部分故障类型及原因

汽轮机通流部分的故障按照发生、发展的进程

分类可以分为突发性故障与渐发性故障^[11]。突发性故障主要包括叶片断裂、阀门杆脱落等。此类故障会让汽轮机系统瞬间发生巨变,产生重大事故。

表1 汽轮机设计参数
Tab.1 Design parameters of the steam turbine

| 项目 | 数值 | 项目 | 数值 |
|-----------------------------|--------|-----------------------------|--------|
| 高压进汽压力/MPa | 10.23 | 低压进汽温度/°C | 304.5 |
| 高压进汽温度/°C | 564.4 | 低压进汽流量/(t·h ⁻¹) | 44.24 |
| 高压进汽流量/(t·h ⁻¹) | 278.63 | 额定功率/MW | 133.01 |
| 中压进汽压力/MPa | 2.221 | ISO工况 ^① 功率/MW | 136.64 |
| 中压进汽温度/°C | 564.2 | 背压/MPa | 0.0049 |
| 中压进汽流量/(t·h ⁻¹) | 305.77 | 冷却水温度/°C | 20 |
| 低压进汽压力/MPa | 0.42 | | |

注:①ISO工况是指环境温度15°C、大气压力0.10135MPa、大气相对湿度60%的工况。

渐发性故障主要包括流道的结垢与磨损。此类故障属于日积月累型的,前期并不会影响汽轮机的运行,只要积累到一定程度时才会引发故障。本文主要研究此类故障。

1.2 汽轮机特征通流面积计算方法

特征通流面积是徐大懋院士提出的无量纲量。其特点为:对于给定的级组,特征通流面积只与几何因素相关而与工况无关;在几何尺寸未发生改变时,无论工况如何的变化,特征通流面积值都不会发生改变^[12-13]。

特征通流面积可由弗留格尔公式变形得到,对于某一级组,设有2个工况(工况1、工况2),各参数应满足弗留格尔公式:

$$\frac{G_1}{G_2} = \sqrt{\frac{p_{01}^2 - p_{z1}^2}{p_{02}^2 - p_{z2}^2}} \sqrt{\frac{T_2}{T_1}} \quad (1)$$

式中, G_1 、 G_2 为变工况前后级组流量, T_1 、 T_2 为变工况前后级组前温度, p_{01} 、 p_{02} 为变工况前后级组前初压, p_{z1} 、 p_{z2} 为变工况前后级组后背压。

等式左右两边变形可得特征通流面积 F :

$$F = \frac{G_1}{p_{01}} \sqrt{\frac{T_1}{1 - \Pi_1^2}} = \frac{G_2}{p_{02}} \sqrt{\frac{T_2}{1 - \Pi_2^2}} \quad (2)$$

式中, $\Pi_1 = p_{z1}/p_{01}$, $\Pi_2 = p_{z2}/p_{02}$ 。

由于弗留格尔公式在推导过程用到了理想气体假设,由公式 $pV = nRT$,可以将 F 中的 T 以 p 和 V 来替代,得到特征通流面积 F_v :

$$F_v = G_1 \sqrt{\frac{V_1}{p_{01}(1 - \Pi_1^2)}} = G_2 \sqrt{\frac{V_2}{p_{02}(1 - \Pi_2^2)}} \quad (3)$$

根据文献[14]中的研究可以发现,公式(2)的精度较低。因此本文采用公式(3)进行计算。

定义特征通流面积退化系数 D :

$$D = F / F_0 \quad (4)$$

式中, F 为汽轮机实际特征通流面积, F_0 为汽轮机初始特征通流面积。

特征通流面积退化系数表征了汽轮机级组通流部分截面积的变化程度。特征通流面积变化系数为 1 代表特征通流面积未发生变化; 特征通流面积退化系数小于 1 代表特征通流面积变小, 对应的级组部分动静叶片结垢、变形和流道受阻等故障形式; 特征通流面积退化系数大于 1 代表特征通流面积变大, 对应级组动静叶片断裂、磨损或腐蚀等故障形式。

1.3 模型验证

分别选取 ISO 100%、75%、60%、50%、40%、30% 工况, 根据电厂提供的热平衡图, 对高中低压缸的特征通流面积进行定量计算, 并建立机组的热力档案, 见表 2。

表 2 不同工况下各级组的特征通流面积
Tab.2 The characteristic flow area of each group under different working conditions

| 级组 | 100% 工况 | 75% 工况 | 60% 工况 | 50% 工况 | 40% 工况 | 30% 工况 |
|-----|------------|-----------|-----------|-----------|-----------|-----------|
| 高压缸 | 90.28 | 90.15 | 90.02 | 90.04 | 90.1 | 90.05 |
| 中压缸 | 205.57 | 204.8 | 203.6 | 202.7 | 201.1 | 199.8 |
| 低压缸 | 598.64 | 602.4 | 608.65 | 615.96 | 620.5 | 627.43 |

由表 2 可以计算得出, 对于高压缸的特征通流面积, 最大误差不过 0.29%, 中压缸最大误差 2.8%, 低压缸最大误差 4.8%, 均小于 5%, 判定此计算模型可用。

2 KNN 算法原理

KNN (K -nearest neighbor) 算法^[15]是一种简单高效的分类算法。其核心思想是: 每个数据样本都对特征空间的 1 个点, 如果在样本空间内 1 个距离当前未知类别数据样本最相邻的 K 个已知类别样本大多数属于某一个类别, 则当前样本也属于这个类别。与其他分类算法不同的是, KNN 算法并不通过判别类域确定样本所属类别, 故 KNN 算法没有传统分类算法的学习过程。其算法简单、计算速度快, 对样本数据集的异常值不敏感^[16-20], 特别适用故障诊断这种多分类问题^[21]。

2.1 传统 KNN 算法

KNN 算法计算方法如式(5)~式(7)所示。首先, 获取已有故障类别标签的性能退化样本数据集:

$$T = \{(X_1, y_1), (X_2, y_2), \dots, (X_N, y_N)\} \quad (5)$$

$$X_i \in \mathbf{R}^n \quad (6)$$

$$y_i \in \{c_1, c_2, \dots, c_k\} \quad (7)$$

式中: X 为故障样本向量的特征值; y 为故障样本标签; c 为故障类别; 上标 n 为故障样本特征数, 即特征向量维度; 下标 i 为故障样本序号, N 为故障样本个数, k 为故障类别数。

通常, 机组采集的蒸汽系统的运行数据种类繁多, 各数据点采集的数据数量级相差较大。直接通过采集的数据样本计算样本之间距离时, 计算结果将会被数量级较大的特征主导, 而数量级较小的特征几乎影响不到计算结果, 故需要通过归一化方法将所有数据映射到同一尺度。

对故障样本特征归一化:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (8)$$

式中, X' 为归一化后的特征值, X_{\max} 为特征在故障样本集中的最大值, X_{\min} 为特征在故障样本集中的最小值。

将故障数据样本映射到同一尺度后, 计算未知类别数据样本到故障样本数据集中各样本的距离。样本距离越远, 表示 2 个样本间的差异越大。根据所研究的分类问题不同, 可以选用不同类型的样本距离度量方式。不同距离度量方式反映了不同意义上的样本差异, 常见的距离度量方式有:

欧式距离 (Euclidean), 表征了空间两点间的真实距离, 计算式为

$$L(x_i, x_j) = \sqrt{\sum_{d=1}^n (x_{id} - x_{jd})^2} \quad (9)$$

曼哈顿距离 (Manhattan), 表征了空间中两点之间的直角边距离, 是对多个维度上的距离求和的结果, 计算式为

$$L(x_i, x_j) = \sum_{d=1}^n |x_{id} - x_{jd}| \quad (10)$$

切比雪夫距离 (Chebyshev), 表征了空间中两点各坐标数值差的最大值, 计算式为

$$L(x_i, x_j) = \max(|x_{i1} - x_{j1}|, |x_{i2} - x_{j2}|, \dots, |x_{in} - x_{jn}|) \quad (11)$$

余弦距离 (Cosine), 表征了空间中两向量夹角余弦值, 反映了数据样本特征向量在方向上的差异, 计算式为

$$L(x_i, x_j) = \frac{\sum_{d=1}^n (x_{id} \times x_{jd})}{\sqrt{\sum_{d=1}^n x_{id}^2} \times \sqrt{\sum_{d=1}^n x_{jd}^2}} \quad (12)$$

式中： L 为样本间的距离； d 为故障样本的第 n 个特征； i 为未知类别数据样本； j 为故障样本集中的样本。

对当前数据样本到故障样本距离排序，选取 K 个距离最小的样本。即这 K 个近邻与当前数据样本最为相似，而当前数据样本的特征值可能具有不同类别的数据样本特征，导致这 K 个近邻并不全部属于同一个类别。此时，通过 K 个近邻的投票来解决这个问题，即 K 个近邻大多数所属的类别，判定为当前数据样本所属类别，故在 KNN 算法中， K 取值通常为奇数，来避免 2 个类别投票结果相同。

2.2 改进 KNN 算法

KNN 算法如图 1 所示。图 1 中，绿色圆形表示当前需要判断类别的数据样本，黑色实线围成的圆圈表示当 $K=5$ 时，当前数据样本的近邻空间。近邻空间内，包含红色方形和橙色三角形 2 个类别的数据样本，其中红色方形有 3 个，橙色三角形有 2 个。按照前述的传统 KNN 算法，根据投票原则，绿色数据样本应当判定为红色方形的类别。而根据图 1，红色方形均处于近邻空间的边缘位置，而橙色三角形距离当前数据样本距离较小，显然，绿色数据样本与橙色三角形数据样本具有更大的相似度，绿色数据样本隶属于橙色三角形类别概率更大。在此种情形下，传统 KNN 算法对数据样本的类别判断失误，究其原因在于传统 KNN 算法投票决策方式认为 K 个近邻在决策过程中具有同等权重，而未考虑到 K 近邻样本距离的大小对决策过程的影响。本文通过样本距离计算 K 近邻样本在决策过程中的权重，评估当前数据样本与 K 近邻样本隶属于同类别的概率，以此对传统 KNN 算法决策过程进行改进。

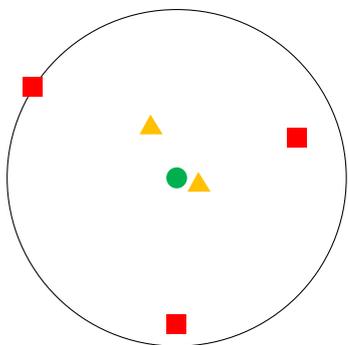


图 1 KNN 算法示意

Fig.1 Schematic diagram of the KNN algorithm

数据样本间距离越小，则认为 2 个样本相似程度越大，故 K 近邻中样本距离小的近邻，在决策过

程中占有更大权重，可以认为权重是样本距离的倒数， K 近邻权重计算方式为

$$W(X_i, X_j) = \frac{1}{L(X_i, X_j)} \quad (13)$$

定义 K 近邻样本距离占 K 近邻样本距离和的比值为当前样本与 K 近邻样本隶属于同一类别的概率。计算当前样本与各 K 近邻样本同故障类型的概率，并对同类别 K 近邻样本概率求和，即

$$P(X_i, X_j) = \frac{W(X_i, X_j)}{W(X_i, X_1) + W(X_i, X_2) + \dots + W(X_i, X_k)} \quad (14)$$

$$P(X_i, C_m) = \sum_{y_k \in C_m} P(X_i, X_k) \quad (15)$$

$$y_k \in C_m \quad (16)$$

式中， $W(X_i, X_j)$ 为 j 样本在 i 样本分类问题中的权重， $P(X_i, X_j)$ 为 i 样本和 j 样本隶属于同一故障类别的概率， $P(X_i, C_m)$ 为 i 样本隶属于第 m 个故障类别的概率。

此时可以实现当前样本相对于各故障类型的不确定度估计，当前样本隶属于各类别的概率表征了当前数据样本隶属于各故障类型的可能性，故隶属概率最大的故障类别即为当前样本所属类别。

2.3 典型设备故障样本建立

本文考虑高中低压缸流道受阻与流道腐蚀 6 种故障型式，具体描述见表 3。

表 3 汽轮机系统典型故障型式
Tab.3 Typical failure modes of steam turbine system

| 故障 | 代号 | 故障描述 | 性能退化系数 |
|---------|----|----------------|------------------|
| 高压缸流道受阻 | 1 | | |
| 中压缸流道受阻 | 2 | 特征通流面积减小，缸效率降低 | $D \in [0.5, 1)$ |
| 低压缸流道受阻 | 3 | | |
| 高压缸流道腐蚀 | 4 | | |
| 中压缸流道腐蚀 | 5 | 特征通流面积增大，缸效率降低 | $D \in (1, 1.5]$ |
| 低压缸流道腐蚀 | 6 | | |

通过汽水系统性能退化模型模拟故障样本，设定汽水系统环境参数为 ISO 工况，余热锅炉入口烟气参数分别设置为 ISO 工况下 100%、75%、50% 负荷时燃机排烟参数，各类故障的设备性能退化系数分别在表 3 预设范围内选取。每种排烟参数下计算 6 种不同性能退化程度数据样本，则每类故障包含 18 个性能退化样本，故障样本总计 108 个。对每类故障，在故障样本参数预设范围内，随机生成 1 个检测样本，其参数设置见表 4。

表4 故障测试样本参数设置
Tab.4 Parameters setting for the failure test sample

| 测试样本 | 性能退化系数 | 环境及燃机排烟参数 |
|------|------------|-----------|
| 1 | $D_n=0.85$ | ISO 100% |
| 2 | $D_i=0.85$ | ISO 75% |
| 3 | $D_i=0.80$ | ISO 50% |
| 4 | $D_n=1.40$ | ISO 50% |
| 5 | $D_i=1.15$ | ISO 75% |
| 6 | $D_i=1.10$ | ISO 50% |

根据上文特征通流面积的计算公式选取高中低压缸的进口蒸汽的温度, 压力和流量作为样本特征。为消除特征参数数量级对样本距离的影响, 通过式(8)对故障样本特征参数归一化。

3 汽轮机故障测试样本验证

3.1 KNN 算法寻优

KNN 算法分类准确性依赖于 K 值和距离度量方式的选取。根据 2.3 节生成的故障样本集对 KNN 算法超参数寻优。对于包含 108 个样本的故障样本集, 随机选取其中 75% 样本作为训练集, 25% 作为测试集, 分别计算 KNN 算法在 $K \in [0, 10]$ 且 $K \in N$, 度量方式分别为欧氏距离、曼哈顿距离、切比雪夫距离和余弦距离的设定下, 通过训练集判断测试集类别的分类正确率。分类正确率计算公式为

$$A_{\text{Accuracy}} = \frac{P_c}{P} \quad (17)$$

式中, A_{Accuracy} 为分类正确率, P_c 为分类正确的样本个数, P 为分类的样本总数。

由于不同训练集划分方式影响测试集上的分类正确率, 故随机多次划分数据样本取分类正确率均值。重复上述过程 1 000 次, 计算分类正确率平均值, 结果如图 2 所示。由图 2 可见: 不同 K 值下, KNN 算法选用余弦距离作为距离度量方式, 正确率均最高; 在选用余弦距离作为距离度量方式时, $K=1, 2, 3$ 分类正确率均高于 0.95, 分类准确性都较为理想; 由于 K 值过小时, 算法分类结果易受数据噪声干扰, 故取 $K=3$, 余弦距离作为距离度量方式。

随机森林算法及 XGBoost 算法是 2 种在复杂分类问题上取得优秀效果的集成算法, 通过建立基于相同故障样本集的随机森林和 XGBoost 算法模型, 与 KNN 算法做对比。随机划分训练集, 通过 2 种算法的最优超参数训练分类算法模型, 并重复 1 000 次, 将 2 种算法平均分类正确率及平均分类时间与 KNN 算法作比较, 结果如图 3 所示。由

图 3 可见, 相较于复杂的集成分类算法, KNN 算法对于本文所述的汽水系统异常设备定位问题具有分类正确率高和分类时间短的优势。

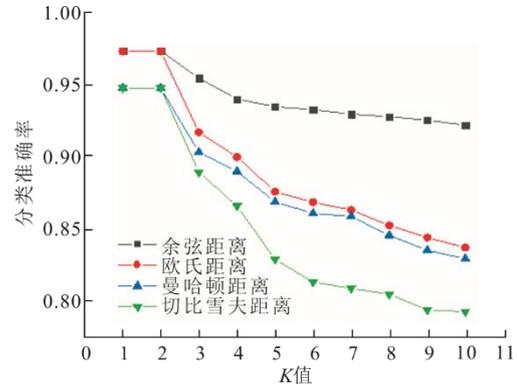


图2 KNN 算法分类正确率随超参数变化关系
Fig.2 Changes of classification accuracy of the KNN algorithm with hyperparameter

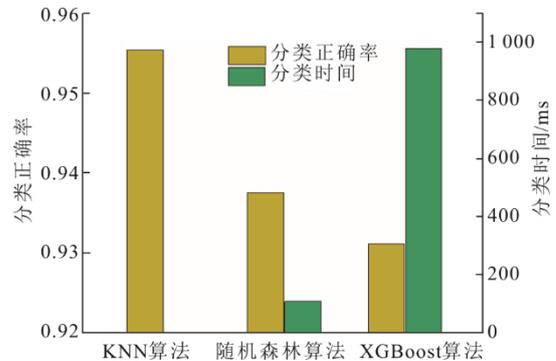


图3 分类算法平均分类正确率与平均分类时间比较
Fig.3 Comparison of average classification accuracy rate and average classification time of classification algorithms

设定 KNN 算法 $K=3$, 余弦距离作为距离度量方式, 前述的包含 108 个样本的故障样本集作为训练集, 对表 3 对应的 6 个故障测试样本的故障类别进行预测。

3.2 传统 KNN 算法的验证结果

首先计算各故障测试样本相对于训练集样本的余弦距离, 找出各故障测试样本余弦距离最小的 3 个近邻, 各故障测试样本的 3 个近邻和余弦距离见表 5。由表 5 可见, 各测试样本距离最小的近邻 1 所属类别均与测试样本的故障类别相符, 近邻 2 中只有 1 个样本所属类别与测试样本故障类别不符, 近邻 3 中有 5 个样本所属类别与测试样本故障类别不符。

通过传统 KNN 算法对故障测试样本所属故障类别进行判别的结果见表 6。由表 6 可见, 6 个测试样本中, 只有测试样本 5 故障类别判断错误, 因

为测试样本 6 的近邻 2 和近邻 3 都属于故障类别 3, 虽然样本距离最小的近邻 1 属于故障类别 5, 但根据投票原则, 测试样本 6 仍被判定为故障类别 3。故传统 KNN 算法的投票决策方式在特定情况下并不能准确利用样本近邻判别当前样本类别。

表 5 故障测试样本近邻与样本距离

Tab.5 The nearest neighbor and distance of fault test sample

| 测试样本 | 近邻 1 | | 近邻 2 | | 近邻 3 | |
|------|------|-------------------------------|------|-------------------------------|------|-------------------------------|
| | 所属类别 | 余弦距离/ ($\times 10^{-3}$) | 所属类别 | 余弦距离/ ($\times 10^{-3}$) | 所属类别 | 余弦距离/ ($\times 10^{-3}$) |
| 1 | 1 | 0.324 | 1 | 0.409 | 5 | 1.726 |
| 2 | 2 | 0.340 | 2 | 0.371 | 3 | 2.191 |
| 3 | 3 | 0.320 | 3 | 0.373 | 2 | 1.576 |
| 4 | 4 | 0.981 | 4 | 1.290 | 4 | 10.134 |
| 5 | 5 | 0.728 | 3 | 0.813 | 3 | 5.495 |
| 6 | 6 | 0.159 | 6 | 0.602 | 1 | 0.669 |

表 6 传统 KNN 算法对故障测试样本的判别结果

Tab.6 The discrimination result of the failure test sample using the conventional KNN algorithm

| 测试样本 | 真实故障类型 | 算法判断故障类型 | 判断结果 |
|------|---------|----------|------|
| 1 | 高压缸流道受阻 | 高压缸流道受阻 | 正确 |
| 2 | 中压缸流道受阻 | 中压缸流道受阻 | 正确 |
| 3 | 低压缸流道受阻 | 低压缸流道受阻 | 正确 |
| 4 | 高压缸流道腐蚀 | 高压缸流道腐蚀 | 正确 |
| 5 | 中压缸流道腐蚀 | 低压缸流道受阻 | 错误 |
| 6 | 低压缸流道腐蚀 | 低压缸流道腐蚀 | 正确 |

3.3 改进 KNN 算法的验证结果

通过改进的 KNN 算法决策方式对测试样本故障类型进行判别, 首先通过式(14)计算测试样本各近邻在分类过程中的权重, 根据式(15)计算测试样本与近邻隶属于同一类别概率, 计算结果见表 7。

表 7 故障测试样本近邻权重和同类别概率

Tab.7 The nearest neighbor weight and the same-class probability of failure test sample

| 测试样本 | 1 | 2 | 3 | 4 | 5 | 6 | |
|------|--------------------------|-------|-------|-------|-------|-------|-------|
| 所属类别 | 1 | 2 | 3 | 4 | 5 | 6 | |
| 近邻 1 | 权重/ ($\times 10^3$) | 3.083 | 2.942 | 3.126 | 1.019 | 2.964 | 6.286 |
| | 同类别 概率 | 0.505 | 0.483 | 0.485 | 0.538 | 0.582 | 0.666 |
| 所属类别 | 1 | 2 | 3 | 4 | 3 | 6 | |
| 近邻 2 | 权重/ ($\times 10^3$) | 2.447 | 2.699 | 2.681 | 0.775 | 1.547 | 1.661 |
| | 同类别 概率 | 0.401 | 0.443 | 0.416 | 0.410 | 0.304 | 0.176 |
| 所属类别 | 5 | 3 | 2 | 4 | 3 | 1 | |
| 近邻 3 | 权重/ ($\times 10^3$) | 0.579 | 0.456 | 0.634 | 0.099 | 0.585 | 1.495 |
| | 同类别 概率 | 0.094 | 0.074 | 0.099 | 0.052 | 0.114 | 0.158 |

样本距离更近的近邻, 在决策过程中被赋予了更大权重, 权重越大, 测试样本与近邻隶属于同一类别的概率越大。测试样本隶属概率最大的故障类别判定为测试样本所属类别, 通过改进型 KNN 算法决策方法判定的测试样本故障类别结果见表 8。

表 8 改进 KNN 算法对故障测试样本的判别结果

Tab.8 The discrimination result of the failure test sample using the improved KNN algorithm

| 测试样本 | 真实故障类型 | 算法判断故障类型 | 判断结果 |
|------|---------|----------|------|
| 1 | 高压缸流道受阻 | 高压缸流道受阻 | 正确 |
| 2 | 中压缸流道受阻 | 中压缸流道受阻 | 正确 |
| 3 | 低压缸流道受阻 | 低压缸流道受阻 | 正确 |
| 4 | 高压缸流道腐蚀 | 高压缸流道腐蚀 | 正确 |
| 5 | 中压缸流道腐蚀 | 中压缸流道腐蚀 | 正确 |
| 6 | 低压缸流道腐蚀 | 低压缸流道腐蚀 | 正确 |

改进 KNN 算法对测试样本集所有样本故障类型均做出了准确判断, 并给出了测试样本隶属于各故障类型的概率。由表 7 可见: 对传统 KNN 算法判断错误的测试样本 5, 改进 KNN 算法计算了此样本 3 个近邻在决策过程中的权重, 分别为 2.964、1.547 和 0.585, 表明近邻 1 在决策过程中起到了主导地位; 通过权重计算样本 5 与各近邻同类别概率, 分别为 0.582、0.304 和 0.114; 而近邻 2 和近邻 3 属于同一类别, 故样本 5 隶属于故障 5 的概率为 0.582, 隶属于故障 1 的概率为 0.418, 最终判断样本 5 隶属于故障 5。

可见改进 KNN 算法通过样本评估近邻在决策过程中的权重, 取得了比传统 KNN 算法更高的分类正确率。并且改进 KNN 算法在本文所讨论的汽轮机系统故障诊断问题上, 取得了良好的效果。

4 结 论

1) $K=3$ 时, 余弦距离作为距离度量方式的 KNN 算法更适用于汽轮机系统故障诊断问题, 其平均分类正确率为 0.955 6。

2) 对测试样本故障诊断结果表明, 改进 KNN 算法比传统 KNN 诊断准确率更高, 对测试样本诊断准确率为 100%, 采用改进 KNN 算法汽轮机系统故障诊断具有可行性。

[参 考 文 献]

[1] 杨志荣. 论我国能源·电力的未来发展[J]. 科学对社会的影响, 2001(2): 35-39.
YANG Zhirong. Future development of energy sources and electric power in China[J]. Impact of Science on Society, 2001(2): 35-39.

- [2] 刘伟, 袁益超, 刘聿拯. 燃气-蒸汽联合循环余热锅炉及其影响因素分析[J]. 电站系统工程, 2008, 24(2): 5-6. LIU Wei, YUAN Yichao, LIU Yuzheng. Gas-steam combined cycle heat recovery steam generator and its influencing factors[J]. Power System Engineering, 2008, 24(2): 5-6.
- [3] 蒋浦宁. 联合循环用汽轮机的开发研制特点[J]. 热力透平, 2003, 32(4): 218-224. JIANG Puning. Development features of combined-cycle steam turbine[J]. Thermal Turbine, 2003, 32(4): 218-224.
- [4] 李帆. 特征通流面积理论在电站汽轮机通流部分故障诊断中的应用研究[D]. 长沙: 长沙理工大学, 2013: 2-9. LI Fan. Application of characteristic flow area theory in fault diagnosis of flow passage of steam turbine in power plant[D]. Changsha: Changsha University of Science and Technology, 2013: 2-9.
- [5] 方名菊. 基于案例的汽轮机智能故障诊断[D]. 杭州: 浙江大学, 2018: 1-17. FANG Mingju. Case-based intelligent fault diagnosis of steam turbine[D]. Hangzhou: Zhejiang University, 2018: 1-17.
- [6] 张红民, 杨利, 魏书军, 等. 汽轮机通流部分故障诊断方法研究[J]. 电工文摘, 2016(5): 57-59. ZHANG Hongmin, YANG Li, WEI Shujun, et al. Research on fault diagnosis method of steam turbine flow part[J]. Electrician Abstracts, 2016(5): 57-59.
- [7] 董晓峰, 顾煜炯, 邱影, 等. 基于聚类分析与加权模糊逻辑的汽轮机组振动故障诊断方法研究[J]. 热能动力工程, 2011, 26(3): 275-279. DONG Xiaofeng, GU Yujiong, QIU Ying, et al. Study on the method of steam turbine unit vibration fault diagnosis based on cluster analysis and weighted fuzzy logic[J]. Journal of Engineering for Thermal Energy and Power, 2011, 26(3): 275-279.
- [8] 董晓峰, 顾煜炯, 杨昆, 等. 汽轮机通流部分故障诊断方法研究[J]. 中国电机工程学报, 2010, 35(2): 71-77. DONG Xiaofeng, GU Yujiong, YANG Kun, et al. Study on fault diagnosis of steam turbine flow path components[J]. Proceedings of the CSEE, 2010, 35(2): 71-77.
- [9] 忻建华, 叶春, 陈汉平, 等. 300 MW 汽轮机高压缸通流部分故障的热参数模糊诊断[J]. 动力工程, 1997(3): 6-9. XIN Jianhua, YE Chun, CHEN Hanping, et al. A study of thermal parameters fuzzy diagnosis for 300MW turbine steam path components failures[J]. Power Engineering, 1997(3): 6-9.
- [10] 郝志莉, 叶春. 电站汽轮机组通流部分故障分析及诊断[J]. 动力工程, 1995, 15(1): 22-26. HAO Zhili, YE Chun. Blading faults of utility steam turbines and their remedies[J]. Power Engineering, 1995, 15(1): 22-26.
- [11] 李帆. 特征通流面积理论在电站汽轮机通流部分故障诊断中的应用研究[D]. 长沙: 长沙理工大学, 2013: 1. LI Fan. Application of characteristic flow area theory in fault diagnosis of flow passage of steam turbine in power station[D]. Changsha: Changsha University of Science and Technology, 2013: 1.
- [12] 薛朝因, 高登攀, 张永海, 等. 基于现场热力参数的汽轮机通流部分故障诊断及应用[J]. 热力发电, 2021, 50(4): 143-148. XUE Chaoyin, GAO Dengpan, ZHANG Yonghai, et al. Fault diagnosis method and application of turbine flow passage based on on-site thermal parameters[J]. Thermal Power Generation, 2021, 50(4): 143-148.
- [13] 张家宁. 大型火电机组通流特性分析及诊断方法研究[D]. 保定: 华北电力大学, 2018: 1-7. ZHANG Jianing. Analysis of flow characteristics and diagnosis methods of large thermal power units[D]. Baoding: North China Electric Power University, 2018: 1-7.
- [14] 徐大懋, 邓德兵, 王世勇, 等. 汽轮机的特征通流面积及弗留格尔公式的改进[J]. 动力工程学报, 2010, 30(7): 473-477. XU Damao, DENG Debing, WANG Shiyong, et al. Application of characteristic flow area of steam turbines and improvement on Flügel formula[J]. Journal of Chinese Society of Power Engineering, 2010, 30(7): 473-477.
- [15] GUO G, WANG H, BELL D A, et al. KNN model-based approach in classification[C]. On the Move to Meaningful Internet Systems: Coopis, Doa, & Odbase-otm Confederated International Conferences, Coopis, Doa, & Odbase, Catania, Sicily, Italy, November. Springer, Berlin, Heidelberg, 2003: 1.
- [16] ZHANG S. Cost-sensitive KNN classification[J]. Neurocomputing, 2019, 391(5): 234-242.
- [17] SONG Y, HUANG J, ZHOU D, et al. IKNN: informative K-nearest neighbor pattern classification[C]. European Conference on Knowledge Discovery in Databases: Pkdd. DBLP, 2007, 4702: 248-264.
- [18] CHEN Z, ZHOU L J, LI X D, et al. The Lao text classification method based on KNN[J]. Procedia Computer Science, 2020, 166: 523-528.
- [19] ZHANG S C, LI X L, ZONG M, et al. Efficient KNN classification with different numbers of nearest neighbors[J]. IEEE Transactions on Neural Networks & Learning Systems, 2018, 29(5): 1774-1785.
- [20] COSTA B G, FREIRE J C A, CAVALCANTE H S, et al. Fault Classification on Transmission Lines Using KNN-DTW[C]. Computational Science and Its Applications: ICCSA 2017. 2017: 174-187.
- [21] ISLAM M M, LEE G, HETTIWATTE S N. Incipient fault diagnosis in power transformers by clustering and adapted KNN[C]. Power Engineering Conference. IEEE, 2016: 1.

(责任编辑 刘永强)