

基于 KNN 的水轮机组轴承监测与故障诊断*

杨铭轩,于亚雄,李 青,邱小波,徐开炜

(南方电网调峰调频发电有限公司检修试验分公司, 广州 510430)

摘 要:为了对水轮机组轴承进行实时监测并对其故障进行诊断,提出一种基于 KNN 算法的水轮机组轴承的故障诊断算法与监测系统。首先,介绍了传统的 KNN 算法,其次引入故障识别球,对 KNN 算法进行优化,使其更适用于水轮机组轴承故障诊断,并对故障样本数据集进行增强和优化;然后,根据水轮机组轴承运行状态能够通过在线油液特征来反映,构建在线油液特征检测系统;最后,通过故障样本数据集对优化后的算法进行验证,并使用设计的监测系统对水轮机组轴承的运行状态进行了一个月的实时监测。结果表明:优化后的 KNN 算法具有可行性与自适应性,对故障识别诊断的准确率高达 0.98;监测系统运行稳定,能够对水轮机轴承进行实时监测,及时对异常数据进行获取与识别,并给出故障诊断报告与检修意见。

关键词:KNN 算法;水轮机组轴承;实时监测;故障诊断

中图分类号:TP277 **文献标识码:**A **DOI 编码:**10.14016/j.cnki.1001-9227.2023.04.066

Bearing monitoring and fault diagnosis of turbine units based on KNN

YANG Mingxuan, YU Yaxiong, LI Qing, QIU Xiaobo, XU Kaiwei

(Generation Company of China Southern Power Grid Co., Ltd Guangzhou 510430, China)

Abstract: In order to monitor the bearings of hydro turbines in real time and diagnose their faults, this paper proposes a fault diagnosis algorithm and monitoring system for bearings of hydro turbines based on KNN algorithm. Firstly, the traditional KNN algorithm is introduced, and secondly, the fault identification ball is introduced to optimize the KNN algorithm, making it more suitable for bearing fault diagnosis of hydro turbine units, and enhancing and optimizing the fault sample data set. Then, according to the operating status of the bearing of the turbine unit, the online oil characteristic detection system can be built by reflecting the online oil characteristics. Finally, the optimized algorithm is verified by the fault sample data set, and the operating status of the bearings of the hydro turbine unit is monitored in real time for one month by using the monitoring system designed in this paper. The results show that the optimized KNN algorithm is feasible and adaptive, and the accuracy of fault identification and diagnosis is as high as 0.98; the monitoring system runs stably, can monitor the turbine bearing in real time, obtain and identify abnormal data in time, and give fault diagnosis report and maintenance opinion.

Key words: KNN algorithm; Hydro turbine unit bearings; Real-time monitoring; Troubleshooting

0 引言

水能资源是现代人类赖以生存的重要资源之一,安全有效地利用水能资源并提供充足稳定的能源,是现代经济发展与人们生活稳定的重要保障,因此,对于水力发电中重要设备水轮机组的运行状态监测,成为社会研究的热点。刘苏程通过对水轮机转轮进行监测,提高转轮使用寿命,从而使水轮机运行更加稳定^[1];冯伟对油膜厚度进行实时监测,来推断水轮机组轴承的运行状态并验证了其方法的可靠性^[2];王忠强对水轮机水利振动的评估手段进行研究,为推动水轮机稳定运行提供了数

据参考^[3]。综上可知,要维持水轮机组轴承的正常运行,需要从多方面对水轮机组轴承进行实时监测并及时对异常数据进行识别与诊断。

1 KNN 算法的基本原理

KNN 算法是一种分类算法,具有计算方法简单、计算速度快、高效等特点,被广泛应用于故障诊断领域。其基本原理是对需要进行测试的样本集进行扫描,随后寻找与该样本集相似度最高的训练样本,最后根据这个样本的类别对测试样本进行分类;同时可以根据样本中不同类别的样本数量分布来进行估计,获取每个测试样本对应每个分类的概率。KNN 算法的基本要素包括分类决策规则、距离度量与 K 值的选择^[4-5]。

根据水轮机组轴承的工作原理获取原始数据样本并进行扫描,将其作为 KNN 算法的特征输入。则 KNN 算法公式可以表示为:

$$T = \{(X_1, y_1), (X_2, y_2), \dots, (X_N, y_N)\} \quad (1)$$

$$X_i \in R^n \quad (2)$$

收稿日期:2022-12-12

* 基金项目:2019 年南方电网调峰调频发电有限公司检修试验分公司科技项目《基于云计算的可定制专家策略和机器学习融合应用的监控数据分析与处理技术平台研发》(020000 KK52180009)

作者简介:杨铭轩(1987-)男,陕西西安人,高级工程师,工程师,主要研究方向为抽水蓄能机组智能监控及大数据分析等。

$$y_i \in \{c_1, c_2, \dots, c_k\} \quad (3)$$

式中, X 代表数据集中故障样本向量的特征值, y 代表故障样本的标签, N 代表数据集中故障样本的个数, 即 $N = 1, 2, \dots, N$, i 代表故障样本的标注序号, k 代表故障分类的个数, c 代表故障的某一类别, n 代表故障样本的特征向量维度。

故障样本的特征进行归一化, 其公式可以表示为:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (4)$$

式中, X' 代表所有数据样本特征归一化后的值, X_{\max} 和 X_{\min} 分别表示样本特征在故障样本集中的最大值和最小值。

1.1 距离度量方式

根据需要研究的分类问题的不同特征, 引入合适的距离度量方式对未知样本与已知样本之间的距离进行计算, 通过样本之间的距离判断样本之间的差异, 从而对数据样本进行分类^[6]。

①计算空间中两个标记点之间的真实距离, 可以选择欧氏距离, 其计算公式如下:

$$L(x_i, x_j) = \sqrt{\sum_{d=1}^n (x_{id} - x_{jd})^2} \quad (5)$$

②计算空间中两个标记点之间的直角边距离, 可以选择曼哈顿距离, 其计算公式如下:

$$L(x_i, x_j) = \sum_{d=1}^n |x_{id} - x_{jd}| \quad (6)$$

③计算空间中两个标记点各自坐标数值差的最大值, 可以选择切比雪夫距离, 其计算公式如下:

$$L(x_i, x_j) = \max(|x_{i1} - x_{j1}|, |x_{i2} - x_{j2}|, \dots, |x_{in} - x_{jn}|) \quad (7)$$

④计算空间中两个标记点的特征向量在方向上的差异, 可以选择余弦距离, 对其夹角余弦值进行计算, 其计算公式如下:

$$L(x_i, x_j) = \frac{\sum_{d=1}^n (x_{id} \times x_{jd})}{\sqrt{\sum_{d=1}^n x_{id}^2} \times \sqrt{\sum_{d=1}^n x_{jd}^2}} \quad (8)$$

1.2 K 值的选择

通过距离度量方式, 对测试数据样本与故障样本之间的距离进行计算, 并根据结算结果进行排序, 随后从中选取距离最短的一个 K 样本, 其中 K 取奇数, 表明测试数据样本与 K 个相邻的故障样本特征最为相似, 同时通过这 K 个样本进行投票, 判断测试数据样本是否属于该类别。

2 基于优化后 KNN 算法的故障样本数据训练

传统的 KNN 算法在分类过程中, 需要不断地对新数据样本和已知故障样本之间的距离进行计算, 而对水轮机组轴承检测过程中, 获取的新样本数据会不断增加, 这导致 KNN 的计算量也不断增大, 且对于 K 的取值, 也更难以确定, 最终致使算法对故障的识别速度变慢以及诊断准确率降低^[7-8]。针对这一问题, 引入一种设定故障识别球的方法, 对 KNN 算法进行优化构建故障诊断系统。

2.1 KNN 算法优化

如图 1 所示, 建立一个数据集 $A = \{a_i\}$, 其中包含无数正常数据样本和各种故障数据样本的检测特征, 并将每一个特征看作是一个故障识别球, 即数据集中有多少种类别的故障数据样本, 就有多少个故障识别球, 以保证数据集的多样性。此时, 新获取样本数据只需要与各个识别球进行对比, 通过新样本与故障识别球之间的相似度来判断其所属类别, 大幅度提高了传统 KNN 的计算速度。

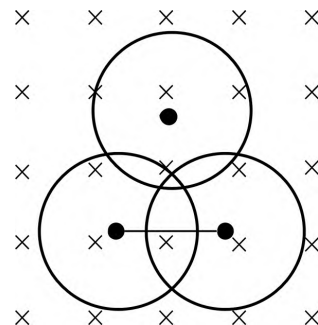


图 1 不同故障识别球在空间中的表达效果

同一类故障识别球之间的欧氏距离不能过小, 以免对故障的识别诊断出现冗余的过程, 影响诊断速度, 同时, 同类别的故障识别球不能太少, 以保证所有的故障能够被多个故障识别球同时识别, 从而提高识别诊断的准确率^[9]。分析图 1 可知, 图中的 \cdot 符号代表着数据集 A 所包含的某一类故障识别球, x 代表该类的故障数据, 而圆则代表故障识别球能够对故障数据进行识别的范围。

2.2 基于优化后 KNN 算法的数据训练

根据水轮机组轴承的工作原理, 收集并建立数据集, 采用优化后的 KNN 算法, 对故障进行识别诊断, 整个流程步骤如下:

(1) 将数据集分为随机集合 A 、训练集 D 以及空数据集 M , 其中, 集合 A 中包含 $s \times t$ 个故障识别球, 即 S 个故障类别, 以及每个类别有 t 个故障识别球; 训练集 D 包含了数据集中除了集合 A 的所有数据样本。

(2) 计算训练集 D 中每一个数据样本特征与集合 A 中所有故障识别球之间的相似度并进行排序, 并将相似度最高的故障识别球单独构建一个集合 B 。

(3) 对集合 B 中的数据进行扩充与增强:

①对集合 B 中相似度高的数据样本进行克隆操作, 从而增加优良数据样本的数量, 最终克隆数量的确定可以通过公式(9)进行计算:

$$n_b = \frac{ar(b, f_i)}{\sum_{k=1}^{|B|} r(b_k, f_i)} \quad (9)$$

②对集合 B 中相似度较低的故障识别球进行变异操作, 并将变异后的故障识别球进行优化增强, 从而获取更多的优良数据样本, 其计算公式如下:

$$c = c \pm (1 - r(c, f_i)) (c - f_i) \quad (10)$$

③通过克隆与变异操作, 扩充了集合 B 中数据样本数量, 这些故障识别球的质量不一, 因此, 需要通过选择操作对故障识别球进行择优选取。

(4) 去除冗余, 对集合 B 中的故障识别球进一步优化, 即先随机删除一个相似度已达标的故障识别球, 同

时对相似度最低的故障识别球再一次进行步骤(3)的操作,对数据样本在此进行扩充与增强,获取新的数据集,随后再次重复去除冗余的步骤,直到所有故障识别球之间的相似度都已达标,获取最优集合 B ,最后将其并入空数据集 M 。

(5)从数据集中重新一组训练数据,进行步骤(2)~(4)的训练,直到所有数据训练完成。

(6)将并入所有最优数据样本的集合 M ,用步骤(4)的方法进行优化,并归入集合 A ,返回步骤(2)进行再次训练,直到迭代次数达到规定数值。

优化后的 KNN 算法在对数据的不断训练过程中,能够及时对系统检测的新故障进行训练,使集合 A 能够不断进行更新扩充,具备自适应的特点。

3 水轮机组轴承监测系统构建

3.1 水轮机组轴承监测特征提取

根据水轮机组轴承的工作原理可知,其运行状态可以通过油液监测特征来反映,如油液中含水量增加,可以推测是水轮机组轴承中密封件出现了损毁;油液中铁磁颗粒增加,可以推测是水轮机组轴承的负载荷过大或

者转速过低等问题^[10-11]。因此,通过对轴承油液的情况进行实时监测,以实现对水轮机组轴承运行状态的监测。

3.1.1 在线油液监测特征提取

在水轮机组运行过程中,实时等间隔地对油液参数进行收集记录,将其变量进行向量化处理^[12],即:

$$f_i = [x_1(t), x_2(t), \dots, x_m(t)] \tag{11}$$

式中, t 代表轴承运行过程中的任一时刻, m 代表监测参数的个数, $x_i(t)$ 代表 t 时刻第 i 个监测参数的变量值, f_i 代表在 t 时刻时的油液监测特征。

对原始数据集的油液监测参数进行特征提取,并将参数特征转化为如下形式:

$$W'_i = [\max(W_i), \min(W_i), \text{mean}(W_i), Q25(W_i), Q50(W_i), Q75(W_i)] \tag{12}$$

特征提取并转化完成后,使用公式(5)计算两个特征之间的距离,对每一个特征进行分类。

3.2 油液监测系统的构建

如图2所示,水轮机组轴承监测与故障诊断系统主要由获取数据的传感器^[13]、数据库以及对数据参数进行检测、预警、故障诊断等功能。

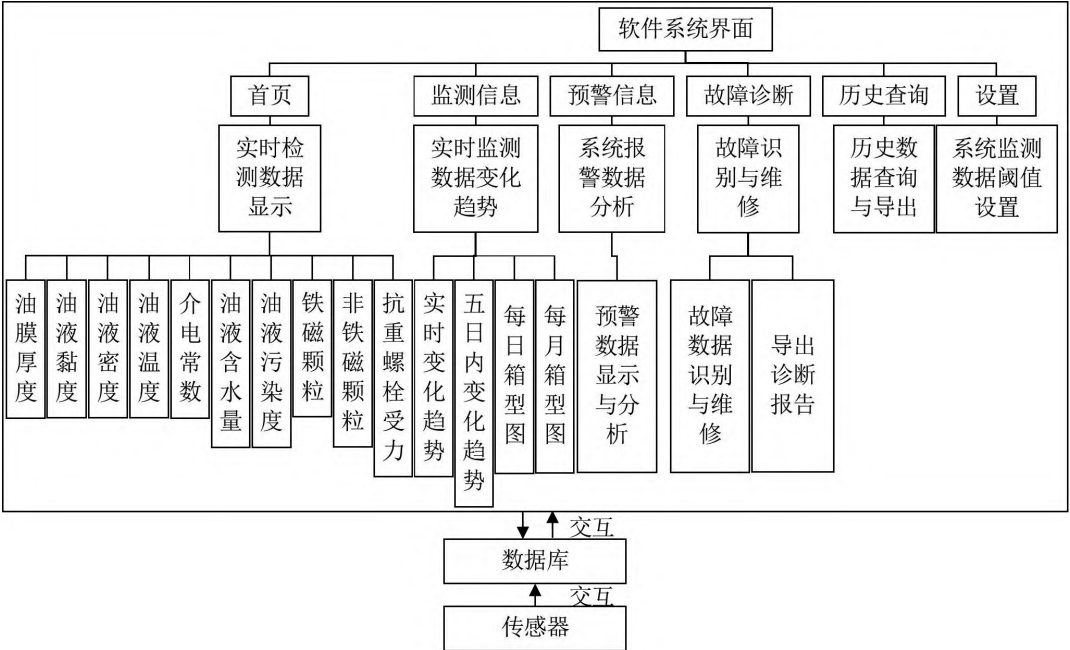


图2 水轮机组轴承监测系统功能模块

其工作流程步骤如下:

(1)在水轮机组轴承运行过程中,采用在线油液数据采集技术,获取各个时间段的油液数据以及参数,通过传感器传输到数据库^[14]。

(2)将所有数据参数在数据库中进行处理,并与系统软件界面进行交互,将处理过程与结果呈现在软件界面,即:

①将油液参数的实时数据,如油液的粘度、密度、温度等参数进行采集和处理后,通过系统软件界面进行显示,以便工作人员对油液的情况进行实时观测^[15];

②根据获取的检测数据,进行特征提取,并生成相应的变化曲线,以便工作人员掌握油液与轴承运行的变化规律与趋势;

③对所有数据进行故障检测,并通过软件界面及时发出预警;

④通过优化后的 KNN 算法对数据进行故障诊断,并在软件界面给出诊断结果与维修建议;

⑤保留历史数据信息并对其进行分析,同时在软件界面导出结果,以便工作人员随时查询与观测其变化。

4 实验结果与分析

4.1 实验数据与实验环境搭建

根据对水轮机组轴承运行机理,对其油液、密封零件、冷却器等运行组件的运行数据进行采集和处理,共获取 500 条原始数据,其中训练集 300 条,测试集 200 条,对优化后的 KNN 算法进行训练和测试验证。

选择 Python+QtCreator 作为设计软件,Python 作为数据处理软件以及各项功能测试开发软件,QtCreator 作为软件界面,采用 MySQL 对数据库进行管理和储存。

4.2 KNN 参数设置

根据水轮机组轴承运行特征,选择粘度、水分、油温、各尺寸铁磁与非铁磁颗粒数以及介电常数作为数据特征参数,将初始化集合 A 中,故障类别设置为 5,即 $s=5$;每个类别的故障识别球设置为 5,即 $t=5$;迭代次数设置为 100,即 $N=50$;相似度最大的故障识别球设置为 8 个,即 $m=8$;克隆系数设置为 10,即 $\alpha=10$;选择比例系数设置为 60,即 $\beta=60$;优化系数设置为 0.8,即 $\varepsilon=0.8$;相似度匹配阈值设置为 0.6,即 $\eta=0.6$ 。

4.3 系统模块划分

在软件系统界面,为了便于工人人员对水轮机组轴承的实时运行状态进行全面的检测,需要将系统根据不同的功能与作用,将所有数据分为不同的模块。

(1) 上机位显示模块

对于油液的在线检测数据,是观测水轮机组轴承运行状态的依据,因此,对监测数据进行实时显示是监测系统设计的核心。

(2) 数据分析及显示模块

为了对水轮机组轴承的运行状态进行全方位的监测,需要将每日的数据进行存档,并以每周、每月单位将数据绘制成折线图,直观清晰地将监测数据的变化趋势显示出来,便于工作人员实时了解水轮机组轴承运行过程中的状态变化。

(3) 故障检测与报警模块

将检测的异常数据进行储存,并发出预警。此模块能第一时间引起工作人员的注意,同时,对异常数据的检测获取,是下一步对故障进行诊断的前提。

(4) 故障的诊断与维修建议模块

在预警模块提供了异常数据后,根据优化后的 KNN 算法,对故障进行识别与诊断,并根据故障的类别与特征给工作提出检查的方向与维修的建议。

(5) 数据查询与导出模块

实时监测的目的之一,是为了通过不同时间的数据对比,来推断水轮机组轴承的运行状态是否稳定,因此,数据储存极为关键,同时需要构建一个模块,使工作人员能够对历史数据进行查询与导出,并通过过往数据的对比,对水轮机组轴承的运行状态进行研究与分析。

(6) 特征向量取值设置

由于用户单位的不同,水轮机组的型号也大小不一,因此,系统的监测特征变量需要根据不同型号水轮机组的特点进行调整,此模块便于工作人员对其进行设置,从而更好地对水轮机组轴承的运行状态进行监测。

4.4 实验结果与分析

4.4.1 KNN 优化算法实验结果与分析

使用 4.2 设置的参数,通过 300 条训练集对优化后的 KNN 算法进行训练,然后通过测试集分别对传统 KNN 算法文提出的优化后 KNN 算法进行测试,则在不同的 K 值下,两种算法在 50 次迭代中的变化曲线图如下:

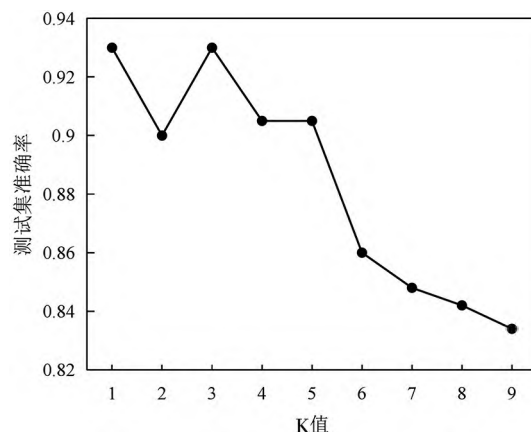


图3 传统 KNN 算法测试结果

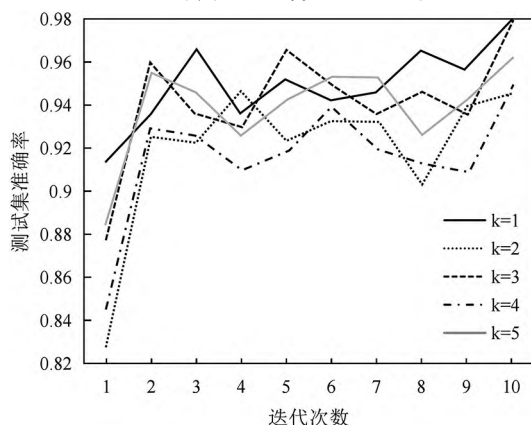


图4 优化后 KNN 算法测试结果

分析图 3 与图 4 可知,传统的 KNN 算法对 200 条测试数据的诊断准确率,最高为 0.93,并且随着 K 值的增大而下跌,最低只有 0.83;而优化后的 KNN 算法对 200 条测试数据的诊断准确率最高可达 0.98,并且随着 K 的增加,还有继续提高的趋势。综合可知,优化后的 KNN 算法具有可行性,并且对故障诊断的准确率大幅度提高,同时因为经过特征提取以及优化后的集合 A 比原始数据集小,其运算速度也有显著提升。

确定优化后的 KNN 算法对故障诊断效果更佳后,建立新故障测试集,再次对算法进行测试,计算新故障数据样本与优化后的故障识别球之间的相似度最大值,其计算结果如下:

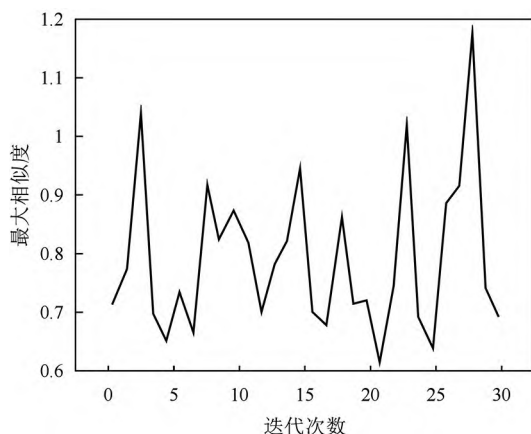


图5 新故障数据样本与所有故障识别球的相似度最大值

分析图 5 可知,所有相似度最大值都超过了相似度匹配阈值,即均大于 0.6,证明优化后的 KNN 算法能够

对新故障进行识别,并且具有一定的稳定性和可靠性,能实现实时监测过程中的动态变化。

综上可得,优化后的 KNN 算法适用于水轮机组轴承的监测与故障诊断,其能够根据实时监测进行动态变化,对故障诊断的准确率高达 0.98。

4.4.2 系统实验结果与分析

采用搭建完成的监测系统,对某一处于运行状态中的水轮机组轴承进行一个月的监测与故障诊断,其实验结果如下:

监测显示界面的模块分区清晰明了,响应迅速,能对油液的实时监测数据进行显示,如图 6 所示:

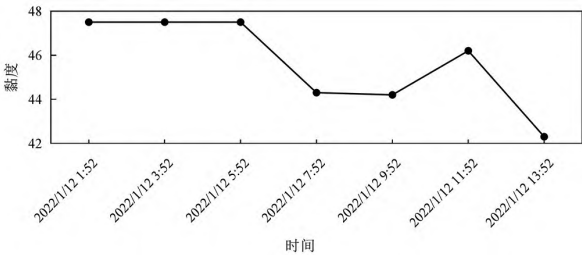


图 6 2022-01-12 油液粘度实时监测数据

监测系统能够详细准确地储存每日数据,并通过折线图在数据分析模块,将每周或每月的数据变化趋势展示出来,如图 7 所示:

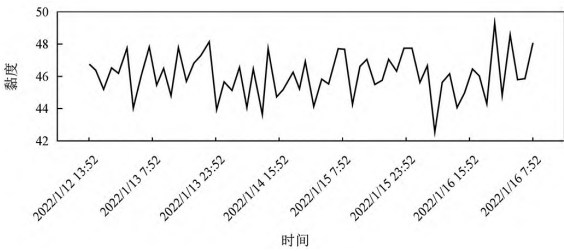


图 7 2022-01-12 油液粘度实时监测数据

监测系统中预警模块能够对异常数据进行获取,便建立数据集传入数据库,故障诊断模块对异常数据的识别与对故障的诊断,准确率保持在 98%左右,并能给出准确的诊断报告,如表 1 所示。

表 1 历史查询界面与设置模块运行正常

监测特征	油液水分含量(ppm)
监测值	110.21
有功功率(MW)	700
水轮机组轴承 转速(rpm)	75
诊断结果	油液中水分含量过高,水轮机组轴承运行不稳定
检查方向	外界水环境是否被污染,机组密封性是否正常
维修意见	对机组零件进行清理并重新密封

5 结论

为了对水轮机组轴承的运行状态进行实时监测与故障诊断,提出一种基于 KNN 的水轮机轴承故障诊断算法以及监测系统。首先,阐述了传统 KNN 的基本原理,其次,根据水轮机轴承的工作原理,对 KNN 算法进行优化,使其更适用于对水轮机轴承的故障诊断,同时,

对水轮机组轴承的故障样本数据集进行增强与优化;然后通过对在线油液特征的提取,搭建水轮机组轴承运行状态监测系统;最后,使用故障样本数据集对优化后的算法进行验证,同时搭建系统测试环境,使用设计的监测系统对水轮机组轴承的运行状态进行一个月的实时监测。结果表明:优化后的 KNN 算法具有可行性,对故障的识别与诊断准确率高达 0.98;监测系统运行稳定,能够实时显示在线油液监测数据,并对其进行储存与分析,同时在优化后 KNN 算法的支持下,能准确地对异常数据进行获取与识别,并提供诊断报告与检修意见。

参考文献

[1] 刘苏程. 基于机器学习水轮机汽蚀自动监测研究[J]. 广东水利水电, 2022, (8): 70-73+88.

[2] 冯伟,李美威,贺石中,等. 大型水轮机组推力轴承油膜厚度在线监测研究[J]. 工程设计学报, 2019, 26(6): 652-657.

[3] 王忠强,张延智,屈伟强. 小浪底水电站水轮机水力振动监测与评估[J]. 人民黄河, 2019, 41(11): 140-143.

[4] 王小霞,欧阳露,郑诗琪,等. GeoHash 与 KNN 在共享单车停靠点优化选择中的应用[J]. 广东工业大学学报, 2022, 39(3): 1-7.

[5] 李磊. 基于 KNN-LSTM 短时交通车流量预测方法的研究[J]. 现代信息科技, 2022, 6(10): 169-173.

[5] 李晨,李艳颖,柴政,等. 基于相互近邻的加权 k 最近邻算法[J]. 宝鸡文理学院学报(自然科学版), 2022, 42(2): 15-22.

[6] 季静静. 基于 TCM-KNN 算法的网络入侵检测方法[J]. 网络安全和信息化, 2022, (7): 144-148.

[7] 杨波,黄倩,付强,等. 基于 CEEMD 和优化 KNN 的离心泵故障诊断方法[J]. 机电工程, 2022, 39(11): 1502-1509.

[8] 郭凯,艾菊梅. 基于 FLANN 改进的 KNN 医疗分类算法[J]. 计算机与现代化, 2022, (8): 25-29+35.

[9] 吴浩然,吴红兰,等. 基于改进 KNN 算法的飞行员工作负荷评估[J]. 航空计算技术, 2022, 52(5): 77-81.

[10] 蔡靖,袁守国,李锐,等. 基于加权 KNN 算法的脑电信号情绪识别[J]. 电子技术应用, 2022, 48(10): 25-30+35.

[11] 陈阳阳,舒胜文,吴涵,等. 金属氧化物避雷器缺陷诊断的反距离加权改进 KNN 算法[J]. 福州大学学报(自然科学版), 2022, 50(5): 635-641.

[12] 贺炳伟. 煤矿奥灰水动态监测系统设计与应用[J]. 陕西煤炭, 2022, 41(6): 183-186+210.

[13] 朱成杰,周亚. 物联网技术下电能质量监测系统的设计[J]. 物联网技术, 2022, 12(11): 30-32.

[14] 魏协奔. 轧机齿轮箱多通道振动监测系统设计与研究[J]. 机床与液压, 2022, 50(21): 194-198.

[15] 田丰,邓晓平,张桂青,等. 基于改进 kNN 算法与暂稳态特征的非侵入式负荷监测方法[J]. 计算机与现代化, 2022, (10): 29-35.

[16] YOU Yi, LIANG Zhe. Non-invasive Load Identification Based on Optimized Extreme Learning Machine[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022, 39(2): 24-29.