

基于不确定性度量的抽水蓄能电站试验检测档案数据识别

岳媛, 王顺军, 洪小青, 朱梦月, 黄力遥

(山东潍坊抽水蓄能有限公司, 山东 潍坊 261000)

摘要: 由于当前抽水蓄能电站试验检测档案数据识别多采用深度信念网络, 缺少对数据特征不确定性的分析, 由此导致其识别准确度较低。为此, 提出基于不确定性度量的抽水蓄能电站试验检测档案数据识别方法。首先, 根据抽水蓄能电站的实际情况, 选取智能识别指标进行数据识别。其次, 采用主成分分析算法获取数据识别综合指标, 结合词嵌入技术提取检测档案数据特征, 对其进行降维处理。最后, 采用偏置矩阵对不确定模型进行优化, 利用新的模型对其进行不确定性度量, 确定数据类别, 完成数据识别。实验结果表明, 在不同样本规模条件下, 应用所提方法对集合中的数据进行识别, 得到的约登指数均优于对比方法, 且当测试样本数量为200时, 约登指数最高可达到85以上, 说明所提方法能有效识别电站试验检测档案的数据类别, 识别精度较高。

关键词: 不确定性度量模型; 抽水蓄能电站; 试验检测档案; 数据识别

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1003-7241(2025)07-0047-05

Identification of Pumped Storage Power Plant Test Archive Data Based on Uncertainty Measurement

YUE Yuan, WANG Shunjun, HONG Xiaoqing, ZHU Mengyue, HUANG Liyao

(Shandong Weifang Pumped Storage Co., Ltd., Weifang 261000, China)

Abstract: At present, the archival data recognition of pumped storage power plant test detection mostly uses deep belief network (DBN), which lacks the analysis of data feature uncertainty, resulting in low recognition accuracy. Therefore, this paper proposes an identification method of pumped storage power station test archive data based on uncertainty measurement. Firstly, according to the actual situation of the pumped storage power station, the intelligent identification index is selected for data identification. Secondly, the principal component analysis algorithm is used to obtain the comprehensive indicators of data recognition, and the word embedding technology is used to extract the features of the detected archive data, and the dimension is reduced. Finally, the bias matrix is used to optimize the uncertainty model, and the new model is used to measure the uncertainty, determine the data category, and complete the data identification. The experimental results show that under different sample sizes, the Youden index obtained by the proposed method is better than that of the comparison methods, and when the number of test samples is 200, the Youden index can reach more than 85, which indicates that the proposed method can effectively identify the data category of the power plant test test files, and the recognition accuracy is high.

Keywords: uncertainty measurement models; pumped storage power stations; test and testing records; data recognition

0 引言

近年来,随着全国经济社会高速发展和电网容量扩大、变电站改造、输配电设备技术提升及对电源规划设计等方面不断提高了要求。国外对于不确定性度量方法研究也比较深入:美国最先采用随机模型来进行风险评估与预报并建立相应数学公式;法国提出利用神经网络构建模糊理论来预测未来可能发生的概率,并且将其应用于电网机组优化决策中去,从而提高了电网调度的可靠性和稳定性。我国电力行业发展起步较晚但发展迅速,有较

多的应用实例。试验检测档案作为抽水蓄能电站运行管理的重要基础资料,其准确性和完整性直接关系到电站的安全性和经济性。然而,传统的数据识别方法往往存在识别效率和准确性低等问题,无法满足抽水蓄能电站对测试检测档案数据进行高效准确识别的需求。因此,研究抽水蓄能电站试验检测档案数据的识别方法具有重要的现实意义和应用价值。

文献[1]中提出的方法利用深度学习模型对雷达数据进行特征提取和分类,在模型训练过程中,对大量数据进行迭代优化,使模型能够学习数据的内在规律和特征,并使用训练后的模型来识别和预测新的数据。该方法可以自动学习数据的特征,无需手动设计和特征提取,大大

*基金项目:国网新源科技项目(SGXYKJ-2021-099)

收稿日期:2024-04-28

地提高了数据处理的自动化水平。然而,深度学习模型的性能在很大程度上取决于训练数据的质量和数量。如果训练数据不足或有偏差,可能会导致模型的识别性能较差。文献[2]中提出的方法通过特定的传感器实时捕捉所需的数据流,使用深度学习算法构建识别模型,并训练该模型自动识别和分类输入数据,可以实现对数据的实时处理和分析,从而及时响应各种需求或变化。但由于需要实时处理大量数据,对计算资源和存储空间的需求很高,导致成本增加。文献[3]中提出的方法通过对数据进行预处理,以提高数据质量和模型的训练效率,通过正向传播计算输出值,并根据输出值与实际标签之间的差异,使用反向传播算法更新网络的权重和偏差,进而实现数据识别。该方法具有较强的泛化能力,可以处理不同领域和场景中的高比例数据识别任务。但神经网络的性能受到各种参数的影响,如网络结构、学习率、正则化参数等。调整过程很复杂。文献[4]中提出的方法根据特定的规则或模式重新排列原始数据,以便突出显示数据中隐藏的结构或特征,对重新排列的数据执行二维FFT(快速傅立叶变换)处理,并通过分析和比较FFT变换后的数据特征,可以实现数据的自动分类和识别。该方法可以提取数据中隐藏的周期模式和频率特征,从而实现准确的识别。但二维FFT变换的计算复杂度相对较高,处理大规模数据集可能需要更长的计算时间。文献[5]中提出的方法通过分析预处理后的数据信号,提取与识别任务相关的关键特征,构建分类器以识别不同类别的数据信号。该方法可以减少单个信号的不稳定性或干扰,提高识别的鲁棒性。但由于不同数据信号的特征和噪声模式不同,预处理和特征提取过程可能更复杂。

本研究以抽水蓄能电站的测试和检测记录为重点,提出了一种基于先进算法的数据识别方法,不仅有助于提高抽水蓄能电站的运行管理水平,也为其他类似领域的数据识别问题提供了有益的参考和指导。

1 抽水蓄能电站试验检测档案数据识别方法设计

1.1 建立智能识别指标

在抽水蓄能电站中,测试和检测档案数据包含大量的设备运行信息、性能参数和故障记录。为了从这些海量的数据中提取有价值的信息,应设置一系列能够准确反映设备状态、性能和故障情况的不确定性识别指标^[6]。通过监测和分析这些指标,数据识别算法可以更准确地识别设备的运行状态和潜在问题。

本文获取待识别抽水蓄能电站试验检测档案数据,选取能够表征电站运行状态的识别指标,具体如表1所示。

表1 电站检测档案数据智能识别指标

指标名称	描述	评价内容
设备运行时间	f_1	用于评估设备的运行时长和老化程度
温度变化范围	f_2	判断是否存在异常温升或散热不良等问题
振动水平	f_3	判断设备是否存在故障或磨损
电气参数	f_4	判断设备的电气性能是否正常
液压参数	f_5	判断液压系统的运行状态是否存在泄漏等问题
油液质量	f_6	分析其清洁度、黏度等指标,以评估油液的性能和更换周期
故障记录次数	f_7	分析故障发生的规律和原因
维修记录	f_8	评估设备的维修质量和维修周期
性能效率	f_9	判断设备的运行效果和经济效益
安全保护性能	f_{10}	检测设备的安全保护装置和措施,如过载保护、短路保护等,确保设备的安全运行

根据抽水蓄能电站的实际情况和需求,选取上述智能识别指标。这些指标将作为数据识别的基础。通过对这些指标的监测和分析,可以实现电站设备的状态监测、故障诊断和性能评估^[7],为电站的安全运行和故障预警提供有力支持。

1.2 试验检测档案数据特征降维

电站试验检测档案数据识别的过程需要分析指标之间的相关关系。通过对原始指标数据进行变换与投影,将高维数据转换为低维数据,同时保留其中的关键信息,进而利用指标的关联性挖掘数据的关键特征,便于后续的数据识别^[8]。

本文采用主成分分析算法对初始指标进行线性组合,以构造新的数据识别综合指标。首先对所有指标进行归一化处理,以统一其物理量纲,表达式为

$$f'_i = \frac{\max f_j - \min f_j}{n \times f_j} \quad (1)$$

式中, $\max f_i$ 、 $\min f_i$ 分别表示指标中的最大值与最小值; n 表示指标个数; f_i 表示第*i*个原始指标。

对指标数据进行统计量检验,获取指标间的相关系数,即:

$$a_{i,j} = f'_i \times \frac{a_k}{\sqrt{A_0/l_p}} \quad (2)$$

式中, $a_{i,j}$ 表示指标间的相关系数; a_k 表示统计量检验的取值范围; A_0 表示平移矩阵; l_p 表示主成分特征值。

利用关联分析法对数据特征因子进行旋转操作,以获取综合指标^[9],表达式为

$$R(f'_i) = \sum \sum \frac{\|a_{i,j} \times \chi_i\|^2}{2 \times \phi_0 \times j_c} \quad (3)$$

式中, $R(f'_i)$ 代表综合指标; χ_i 表示特征贡献率; ϕ 表示评价一致性矩阵; j_c 表示分辨系数。

利用词嵌入技术提取检测档案中数据的文本特征^[10], 表达式为

$$e_1 = \frac{R(f_i) \bullet m_0}{w_k \bullet d(x)} / r_p \quad (4)$$

式中, m_0 表示数据离散程度值; w_k 表示数据属性向量; $d(x)$ 表示数据空间偏移程度值; r_p 表示模拟因子。

引入搜索线性变换矩阵对原始数据和指标进行投影与变换, 以获取两个相邻样本间的马氏距离, 计算公式如下:

$$D_m = \sum_{k=1}^n \frac{e_1 \times \mu_0}{h_e \times b_0} \quad (5)$$

式中, D_m 表示两个相邻样本间的马氏距离; μ_0 表示样本相似概率; h_e 表示第 e 个指标的权重; b_0 表示白化矢量; n_1 表示样本数量。

在样本中随机选出一个近邻样本, 并以欧式距离为标准, 计算 D_m 到少数类样本中所有样本的距离^[11], 表达式为

$$X_d = \frac{D_m(1-p_0)}{g_0 \exp\left(-\frac{\kappa_0}{q_c}\right)} \quad (6)$$

式中, p_0 表示模糊分类矩阵; g_0 表示隶属度矩阵; κ_0 表示决策属性; q_c 表示激活函数。

在上述内容的基础上, 对完成识别指标处理的数据特征进行降维, 表达式如下:

$$u_s = X_d \times \gamma_c \times s_0 / \tau_p \quad (7)$$

式中, γ_c 表示数据类型; s_0 表示原始数据集; τ_p 表示转换系数; u_s 表示降维后的数据特征。

在智能识别指标基础上, 采用主成分分析算法获取数据识别综合指标, 结合词嵌入技术提取检测档案数据特征, 进而对特征进行降维处理, 为接下来实现数据识别提供便利条件。

1.3 基于改进不确定性度量模型的数据识别

通过上述计算与分析, 本文采用改进不确定性度量模型实现对检测档案数据的识别。改进不确定性度量模型能够较好地处理数据中的不确定性和模糊性, 可以保证数据识别的可靠性和稳定性^[12]。基于改进不确定性度量模型的数据识别流程如图1所示。

首先采用偏置矩阵对不确定模型进行优化与改进, 以得到更新后的运算模型^[13], 表达式为

$$M = t_s \times t_s \quad (8)$$

式中, t_s 表示不确定水平的界限; t_s 表示偏置矩阵。

将降维后的数据作为基础数据, 利用新的模型对其进行不确定度量, 以量化每个特征的不确定性程度, 表达式为

$$r_p(M) = \frac{u_s}{\delta_c} \quad (9)$$

式中, u_s 表示数据降维后的特征; δ_c 表示连接权重。

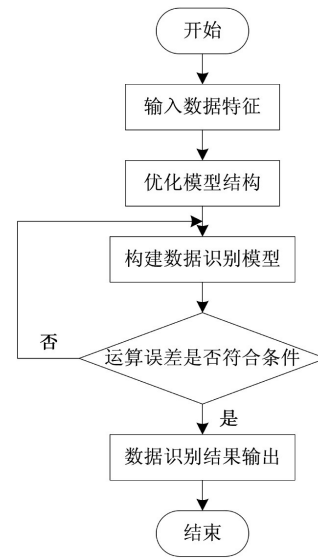


图1 基于改进不确定性度量模型的数据识别流程

进一步构建数据识别模型, 如下式所示:

$$\Omega = r_p(M) \sum \frac{\omega_0}{v_n} \quad (10)$$

式中, ω_0 表示迭代次数; v_n 表示学习率。

若当前算法的运算误差符合预设条件, 则立即输出数据识别结果^[14], 表达式为

$$SPEC = \frac{\Omega(1+v_\theta)}{\zeta_0 \times J(x)} \quad (11)$$

式中, v_θ 表示数据的比例系数; ζ_0 表示电站具体运行参数; $J(x)$ 表示核函数; $SPEC$ 表示数据的特异性程度。

根据数据特异性程度确定数据类别, 用于评估设备的运行情况^[15]。至此, 完成基于改进不确定性度量模型的抽水蓄能电站试验检测档案数据识别方法的设计。

2 实例论证分析

为验证本文设计的改进不确定性度量模型的抽水蓄能电站试验检测档案数据识别方法在实际应用中的效果, 将本文方法应用在某抽水蓄能电站中, 并对其数据进行识别, 根据实验结果分析该方法的识别效果。

2.1 实验准备

某抽水蓄能主要由上水库、下水库、输水系统、发电厂和控制系统组成。其中, 水泵水轮机为S5600可逆式水泵水轮机, 额定转速: 300 r/min, 额定流量 500 m³/s, 额定水头 400 m, 最大出力 1 200 MW; 交流发电机型号为G9000, 额定容量 900 MVA, 额定电压 15 kV, 额定电流 60 kA, 额定转速 300 r/min。

为了确保电站设备的性能和质量, 电站定期对设备进行测试和检查。测试内容包括设备的机械性能、电气性能、安全性能等方面, 并生成试验检测档案, 其内容如表2所示。

表2 电站试验检测档案内容

检测类别	检测设备	检测内容
机械性能检测	水泵水轮机	轴承振动
		密封件检测
	发电机	转动部件检测
		轴承温度
电气性能检测	发电机电气参数	转子动平衡
		绝缘电阻
	继电保护与自动装置	温升
		电压与电流
安全性能检测	消防设施	继电保护装置校验
		自动装置校验
	安全阀与紧急停机装置	消防设备检查
		模拟火灾演练
		安全阀压力测试
		紧急停机装置测试

实验平台采用高性能计算机集群,配备多核处理器和大容量内存,满足数据处理和识别的需求。同时,实验平台还配备了数据采集设备,用于从抽水蓄能电站测试档案中采集数据;数据存储设备,用于存储采集到的实验检测数据,确保数据的完整性和安全性;数据处理软件,用于对采集到的数据进行预处理、清理和格式化,以便进行后续的数据识别和分析。抽水蓄能电站的环境如图2所示。



图2 抽水蓄能电站

使用数据采集设备从抽水蓄能电站的测试档案中采集数据,并将数据存储指定的存储设备中;使用数据处理软件对采集到的数据进行预处理,包括数据清理、格式转换等,以消除数据中的噪声和异常值;利用本文方法识别预处理后的数据,提取设备状态、性能参数等关键信息。

通过上述实验准备和具体的数据描述,可以模拟抽水蓄能电站测试档案数据识别的实验过程,为后续数据处理和分析提供有力支持。

2.2 实验说明

为获得改进不确定性度量模型的控制参数,在采集的样本中随机选取100个数据作为实验数据集,用于分类任务,每个样本包含20个特征。设置初始正则化参数为0.01,最大迭代次数为1000,回归学习率为0.02,置信度阈值为0.7;利用5折交叉验证来评估模型的性能。

实验采用对比的形式进行,选取当前较为成熟的数据识别方法作为对照组方法,本文方法作为实验方法,并

基于实验结果分析不同方法的实际应用性能。

2.3 数据识别结果与分析

基于以上实验准备与相关参数的设置,将100个数据样本依据属性划分为2个类别的数据组,并将本文设计的方法应用于该电站试验检测档案数据识别中,实验结果如图3所示。

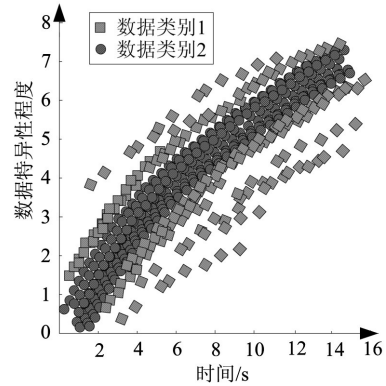


图3 数据识别结果

分析图3可知,应用本文设计的方法对电站试验检测档案进行数据识别,可以实现对数据集中不同类别数据的精准识别,识别结果可以达到预期,表明本文方法的识别应用效果良好。

2.4 对比实验与分析

实验中引入数据挖掘算法(方法1)、模糊规则算法(方法2)作为本文方法的对比方法。分别采用3种方法进行数据识别,采用约登指数作为评估不同方法识别性能的指标。约登指数越高,表明识别准确度越高,识别效果越好。对比结果如图4所示。

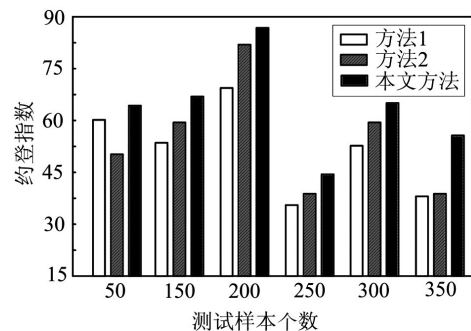


图4 不同识别方法的识别结果对比

根据图4可以看出,在不同样本规模条件下,应用所提方法对集合中的数据进行识别,得到的约登指数显著高于方法1和方法2,说明本文方法的识别效果更加优异。而对照组方法识别精度较低的原因主要是未考虑到数据特征之间相关性的影响,从而使得识别模型易陷入过拟合情况,降低了识别准确度。通过对比实验证明了本文方法的实际应用有效性。

(下转第119页)