DOI:10.19344/j.cnki.issn1671-5276.2024.06.052

IDC 安防巡检机器人网络监控告警系统设计

张晨[1],李志飞[1],陶熠昆[2],张淑兰[2]
(1. 雅砻江流域水电开发有限公司,四川 成都 610051;
2. 浙江国自机器人技术股份有限公司,浙江 杭州 310053)

摘　要:IDC 安防巡检机器人的网络监控系统拓扑结构复杂且存在盲区,以整体思路设计的告警系统可能导致机器人无法准确监测到所有的安全事件或异常情况。提出基于分布式 Hadoop 的 IDC 安防巡检机器人网络监控告警系统设计。引入 HDFS 体系结构及 MapReduce 任务处理技术,设计 Hadoop 高性能监控数据处理平台,降低运维成本;在该平台上设计 IDC 安防巡检机器人网络监控系统,对数据中心实施全面巡检;构建基于能量特征变换的无阈值告警算法,提取监控项数据中的能量、波动以及时间 3 种特征,在此基础上将随机森林算法融合到模型中,自动形成巡检软件设计并触发告警。实验结果表明:所提方法具有较高的告警准确率和效率,且能够有效降低 IDC 运维成本。

关键词:Hadoop 平台;IDC;安防巡检机器人;监控告警;能量特征变换提取;随机森林
中图分类号:TP242.3;TP277　　**文献标志码**:B　　**文章编号**:1671-5276(2024)06-0262-06

Design of Network Monitoring Alarm System for IDC Security Patrol Robot

ZHANG Chen[1], LI Zhifei [1], TAO Yikun[2], ZHANG Shulan [2]
(1. Yalong River Hydropower Development Co., Ltd., Chengdu 610051, China;
2. Zhejiang Guozi Robot Technology Co., Ltd., Hangzhou 310053, China)

Abstract: As the network monitoring system topology of IDC security patrol robot is complex in structure with blind spots, its alarm system may cause the failure in accurately monitoring all security events or abnormalities. The design of IDC security patrol robot Network monitoring alarm system based on distributed Hadoop is proposed. The HDFS architecture and MapReduce task processing technology are introduced to design a Hadoop high-performance monitoring data processing platform for operation and maintenance cost reduction. The IDC security patrol robot Network monitoring system is designed on the platform to implement comprehensive patrol inspection on the data center. A threshold free alarm algorithm based on energy feature transformation is constructed to extract the three characteristics of energy, fluctuation and time in the monitoring item data. On this basis, the random forest algorithm is integrated into the model, automatically forming the patrol software design and trigger the alarm. The experimental results show that the proposed method has high alarm accuracy and efficiency and can effectively reduce IDC operation and maintenance costs.

Keywords: hadoop platform; IDC; security inspection robot; monitoring alarm; energy feature transformation extraction; random forest

0　引言

互联网数据中心(IDC)的运行维护是一项重要且艰难的工作。IDC 可使企业从繁杂、重复性以及低价值的维护工作中解脱出来,投入到价值更高的处理工作中去,虽然极大地提高了生产能力和效益,但随之而来的安全问题也越来越多。数据运营公司在人工智能不断发展的背景下,迫切需要考虑如何引入 IDC 安防巡检机器人并采用完备的有效网络监控系统来应对突发的灾难性事件,对异常情况及时告警并处理,保障网络安全。因此 IDC 安防巡检机器人网络监控系统告警技术已成为该领域的研究重点。

赵庆兵等[1]利用参数自回归算法实现 IDC 巡检机器人网络监控系统的早期预警。利用多维度时序数据进行参数自回归,实现了对系统正常工作状态的估计,并从实测值中提取残差的特征,在此基础上设定一个动态阈值,根据该阈值完成系统的状态监测和告警机制。但方法容易产生误告警问题,加重运维成本。林凌云等[2]利用知识图谱技术建立 IDC 安防巡检网络监控系统的故障告警策略。该方法首先以系统历史事故和变位信号作为 IDC 监控系统事故行为图谱;其次利用异常行为建立系统异常行为图谱;最后从图谱中挖

第一作者简介:张晨(1981—),男,四川成都人,工程师,本科,研究方向为电子信息工程,zdz5412145@yeah.net。

掘出告警信号之间的行为逻辑并显示出来。该方法告警效率较低,同时也降低了系统的运行效率。

为了解决上述系统设计方法中存在的问题,本文提出一种 IDC 安防巡检机器人网络监控告警系统设计方法。

1　系统基础 Hadoop 平台设计

IDC 的海量运维监控数据是动态实时的,频繁的数据采集,降低了监控效率,不能及时得到作业系统状态问题的反馈,且增加了服务器运行成本。Hadoop 平台具有低成本、大规模、高效安全的优势[3],因此将整个 IDC 安防巡检机器人网络监控系统建立在 Hadoop 之上是实现低成本、高性能处理平台的必要方式,其体系结构如图 1 所示。

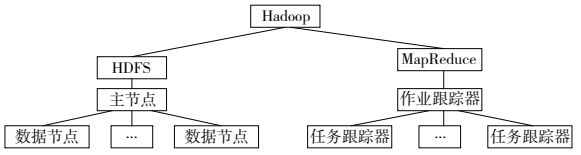

图 1　Hadoop 体系结构

Hadoop 主要由提供数据存储功能的 HDFS 和在集群环境中提供数据分析计算功能的 MapReduce 组成。

1) HDFS

HDFS[4]是一个独立的分布式文件系统,其容错性较高,能够根据 IDC 安防巡检机器人网络监控系统所需要和查询的条件,存储任意形式的数据并自动优化数据流量,保障较高的带宽,降低数据的丢失并提高工作效率。其体系结构如图 2 所示。

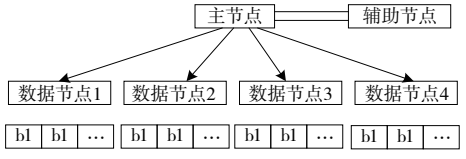

图 2　HDFS 体系结构

2) MapReduce

在 IDC 网络监控数据采集时已经将采集到的数据实时保存在 HDFS 中,基于此可以利用 MapReduce 完成数据的读取和处理。通过 Map 阶段和 Reduce 阶段实现网络监控数据的分析和处理,也同样适用于离线大数据的计算分析,其工作原理如图 3 所示。

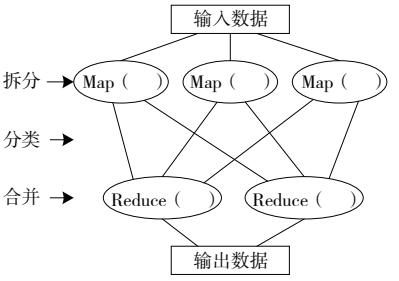

图 3　MapReduce 任务处理原理

由此即可搭建一个数据规模易扩展、处理速度快的分布式海量数据处理监控平台,在该平台上完成后续 IDC 安防巡检机器人网络监控系统告警处理,能够有效降低运行成本和提高安全性。

2　系统的硬件与软件设计

2.1　Hadoop 平台上监控告警系统硬件设计

传统的人工周期性安防巡检存在巡检内容单一、数据重复率较高等缺陷,IDC 运维环境的复杂度以及巡检人员工作能力等因素均会影响安防巡检工作的质量。为此,本文基于 Hadoop 平台,引入 IDC 安防巡检机器人[5]辅助替代人工巡检。IDC 安防巡检机器人的网络监控系统主要由巡检端和上机位监控系统两部分组成,总体设计框架如图 4 所示。

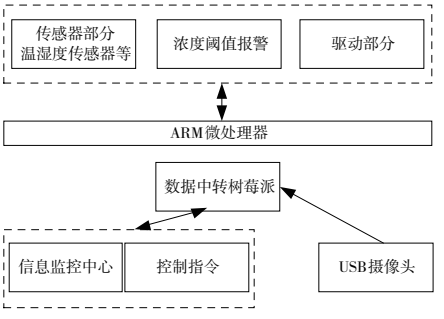

图 4　IDC 安防巡检机器人网络监控系统框架图

巡检端主要负责底层控制端实施数据采集(包含温湿度信息、燃气浓度信息和甲醛浓度信息等)以及自主导航巡检,其结构如图 5 所示。

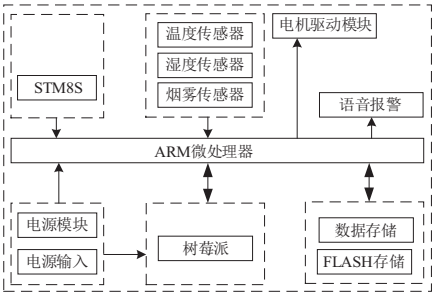

图 5　巡检端框架图

在机器人巡检端中,ARM 微处理器[6]的主要任务是处理采集到的各种数据,若数据超过预警值则执行语音报警和补救,以此减少经济损失,在此基础上利用串口将信息传输给树莓派,再通过树莓派将信息从网络上传到监控系统中。

当 IDC 安防巡检机器人完成巡检任务后,会将采集到的巡检数据信息自动传送到其网络监控系统中,再利用网络监控系统将信息传送给上述设计的 Hadoop 平台中,以此高效率、低成本地完成监控信息存储和分析处理,为后续监控系统告警策略提供有利条件。

2.2　软件算法设计

本文所提方法通过上述方法获取到监控项数据后,建立一种基于能量特征变换提取的无阈值告警算法,该算法能够在满足实时业务告警需求的前提下,有效提高告警准确性,其流程如图 6 所示。

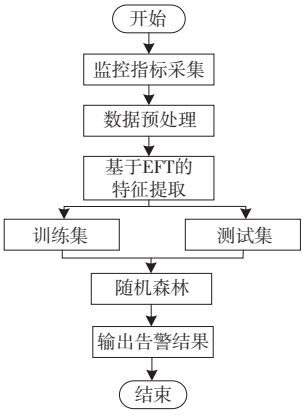

图 6　基于能量特征变换提取的无阈值告警模型流程图

通过能量特征变换方法对监控项数据进行特征提取。能量特征变换是一种基于能量的特征提取方法,它能够捕获到信号中的能量分布特征并提取出能量相关的特征。

1)数据预处理

在监控过程中,由于受到某些主要仪器故障以及巡检环境干扰,会导致在监控数据采集过程中出现部分数据缺失的现象。为了确保监控项数据的质量,利用数据填充法[7],通过调整前后数据点的平均值填补中间的缺失数据,并采用滑窗建立数据集中的样本,示意图如图 7 所示。

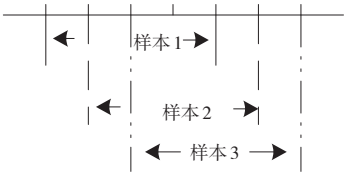

图 7　滑窗构造数据集示意图

在 Hadoop 平台上,滑窗窗口大小预设为 5,监控项每隔 5 min 读取一次监控项数据,经数据预处理后得到 N 个监控值,相应获取(N−5)+1 个数据样本。

2)基于能量特征变换的特征提取

通过数据预处理获取监控数据完整样本后,根据 IDC 监控项数据特点,采用基于能量特征变换的特征提取算法,提取以下 3 个方面的特征值。

a)能量

能量用于反映监控值的大小,提取计算公式如下:

$$\begin{cases} g_1 = u_1 \\ g_2 = u_1 + u_2 + \cdots + u_t \\ g_3 = \sum_{t=1}^{t} u_t^2 \end{cases} \tag{1}$$

式中:g_1、g_2、g_3 分别代表当前值、平均值以及能量值;u 代表监控值;t 代表第 t 个时间点。

b)波动

波动用于反映监控项数据的变化,计算公式如下:

$$\begin{cases} g_4 = (u_t - u_1)/(t-1) \\ g_5 = \sqrt{t^{-1}\sum_{t=1}^{t}(u_t - \eta)^2} \\ g_6 = I \\ g_7 = \sum_{t=1}^{t-1} |u_{t+1} - u_t| \end{cases} \tag{2}$$

式中:g_4、g_5、g_6、g_7 分别表示斜率、标准差、高于均值个数以及相邻绝对波动;η 表示全部监控数据的平均值;I 表示当前窗口内大于 η 的数量。

c)时间特性

时间特性表征当前监控值与时间的关联,计算公式如下:

$$\begin{cases} g_8 = \{1,2,\cdots,7\} \\ g_9 = \{0,24\times60\} \end{cases} \tag{3}$$

式中 g_8、g_9 分别表示以周为周期以及以分钟为周期的关联挖掘时间。将当前窗口 t 和前两个窗口 t−1、t−2 中的特征组合作为最终特征值,表达式如下:

$$A_t = [g_1, g_2, \cdots, g_9] \tag{4}$$

$$G_i = [A_{t-2}, A_{t-1}, A_t] \quad t = 1,2,\cdots,t-3+1 \tag{5}$$

式中:A_t 为每个窗口所提取的特征;G_i 为组合特征集;t 为窗口大小。

3)引入随机森林划分高维监控项数据

利用随机森林来进行高维监控项数据的划

分,输出告警结果。随机森林[8]是一种包含一系列树结构分类器,用$\{j(x,\boldsymbol{\Theta}_k,k=1,2,\cdots)\}$描述,其中的$\boldsymbol{\Theta}_k$表示具有独立均匀分布特性的随机向量,分类过程分为 3 个步骤。

步骤 1:子训练集生成

通过 Bootstrap 方法从大小为 N 的原始监控数据集中采样 N 次,建立 M 个子训练集。

步骤 2:特征选择

从森林中所有树节点中随机选取部分特征,通过基尼指数法选取最优特征划分节点建立决策树[9]。基尼指数越高说明该监控项数据纯度越高,其一致性也越好,计算公式如下:

$$\mathrm{Gini}(F)=G_i-\sum_{k=1}^{K}(|F_k|/F)^2 \tag{6}$$

式中:Gini(·)为基尼指数;F 为给定的数据样本;K 表示其类别;F_k 代表第 k 个类别的数量。若引入特征 A,则需将 F 分为 V_1 和 V_2 两部分,由此得到基于特征 A 的数据样本 F 的基尼指数计算公式:

$$\mathrm{Gini}(V,A)=(|F_1|/F)\mathrm{Gini}(V_1)+(|F_2|/F)\mathrm{Gini}(V_2) \tag{7}$$

步骤 3:由所建立的全部决策树生成 M 颗树组成的森林,每棵树的权重相同,利用森林投票机制选择到最多的类别则是监控项数据样本的最终分类结果[10],流程如图 8 所示。

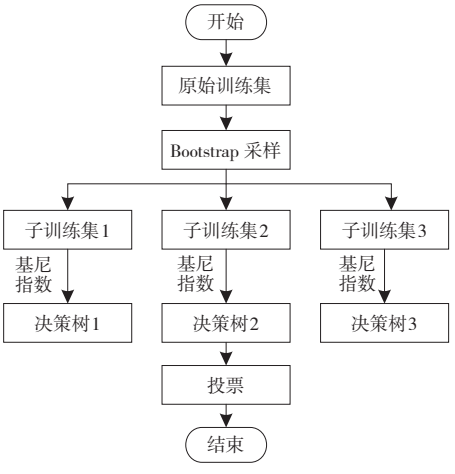

图 8　随机森林算法流程图

随机森林具有较快的训练速度、较高的分类精度以及较强的抗噪能力,能够有效划分高维监控项数据。为此将其引入到无阈值告警模型中完成训练,训练后的结果传输到 Hadoop 平台上的 IDC 安防巡检机器人网络监控系统中,即可根据实际情况实时触发告警并执行相应的措施。

3　实验与分析

3.1　实验设置

为了验证基于 Hadoop 的 IDC 安防巡检机器人网络监控系统告警方法的整体有效性,需要对其展开测试。实验对象为 IDC 数据机房巡检机器人 RW400,在主测计算机上安装 Hadoop 分布式文件系统和 Hadoop 计算框架,搭建以 IDC 安防巡检机器人为核心的机器人网络监控系统告警实验平台,实现相关软件的运行。

IDC 安防巡检机器人的巡检环境设置如下。

1)摄像头布置:实验选择 500 m^2 的中型 IDC 机房,在 IDC 机房内合理地布置摄像头以覆盖关键区域和设备。摄像头可以固定在天花板上或挂在墙壁上,确保能够完整记录需要巡检的区域。

2)传感器安装:在机房内部安装合适的 LM35 温度传感器、DHT22 湿度传感器、MQ-2 烟雾传感器,以便监测环境状态并及时发现异常情况。

实验参数如下。

1) Hadoop 集群的节点数量:共设置 3 个节点,其中一个用作主节点,其余节点用作工作节点,以提供基本的容错能力和高可用性。

2)使用 YARN(yet another resource negotiator)来管理和分配计算资源,根据每个任务的需求动态分配资源。

3)数据分片大小:128 MB 或 64 MB。

3.2　结果分析

采用本文所提方法和文献[1]方法、文献[2]方法开展误警率和告警准确率、稳定度以及运行时间 3 个方面的测试。

1)误警率和告警准确率

在 IDC 运维环境巡检监控告警过程中,IDC 安防巡检机器人极易出现误警情况,导致告警精度较低,为此将误警率作为衡量 IDC 安防巡检机器人网络监控系统告警有效性的重要指标之一。误警率指的是本来正常的数据被误认为异常数据并触发警告,利用所提方法和文献[1]方法、文献[2]方法并展开 12 次测试,得到图 9 的对比结果。

由图 9 可以看出,在实验次数不断增加的前提下,文献[1]方法和文献[2]方法的误警率波动较大,文献[1]方法的误警率呈较大的增长趋势,而文献[2]方法曲线波动较大,稳定性较差;相比之下,所提方法的误警率最小,最多不超过 15%,说明所提方法具有更好的运行性能。

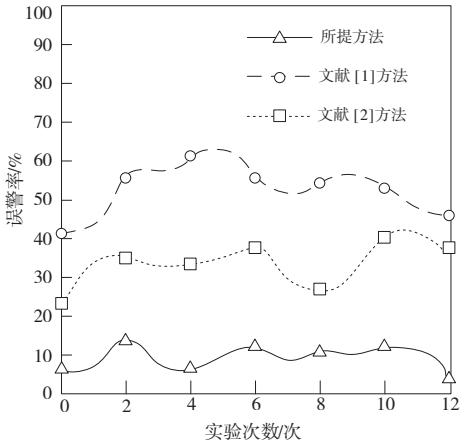

图 9　不同方法误警率对比

告警准确率指的是异常监控数据被成功判断为异常数据的概率以及正常监控数据被判断为正常数据的概率,将其作为评估指标,测试 3 种方法的告警准确率,结果如图 10 所示。

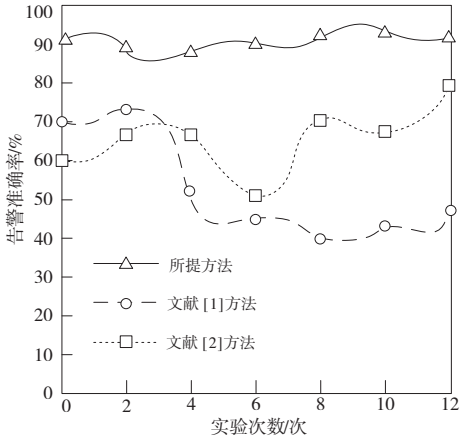

图 10　不同方法告警准确率对比

由图 10 可知,随着实验次数的增加,所提方法的告警准确率始终保持在 88%以上且过程较稳定;文献[1]方法的告警准确率在前 3 次测试中虽然能够保持在 65%左右,但随着实验次数的增加,该方法的告警准确率出现变化,告警准确率最低已经降到 32%,会直接导致误告警的发生;文献[2]方法的告警准确率在全部测试中保持在 50%~80%的范围内。对比可知,所提方法的告警准确率相对较高,能够有效增强网络的安全性。

2)稳定度

告警方法的稳定度是保障 IDC 安防巡检机器人监控性能的第二个重要指标,稳定度越高,系统运行的性能和可靠性越强,随着监控数据的不断增加,3 种方法的稳定度变化情况如图 11 所示。

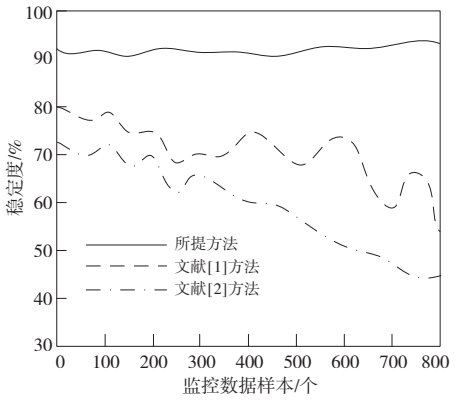

图 11　不同方法稳定度对比

从图 11 中可以看出,其他两种传统方法在监控数据较少的情况下,稳定性较好,但随着监控数据的增加,文献[1]方法的运算稳定度出现大幅度下降的趋势,文献[2]方法的运算稳定度曲线不平稳。由此可以说明所提方法具有更高的运算稳定度,在提高系统运维稳定性的同时,也能相应地降低运行时间。

3)运行时间

在监控数据异常情况下的及时告警是保障 IDC 运维环境安全的关键。告警方法的运行时间越少,表明其告警实时性越高,对运维环境安全的可靠性越强。在与上述测试条件相同的基础上,采用 3 种方法对监控数据分别展开异常告警测试,结果统计如表 1 所示。

表 1　不同方法运行时间对比

监控数据/个	运行时间/ms		
	所提方法	文献[1]方法	文献[2]方法
100	7.88	20.82	10.02
200	8.38	21.45	10.79
300	8.79	22.47	11.23
400	9.25	23.05	11.97
500	9.77	23.98	12.35
600	10.26	24.55	12.97
700	10.85	25.21	13.66
800	11.42	25.98	14.25

分析表 1 可得,所提方法在监控数据增加到最多的 800 个时,运行时间为 11.42 ms,整体平均耗时为 9.58 ms;而文献[1]方法耗时最长,平均为 23.44 ms,高于所提方法 13.86 ms;文献[2]方法虽然低于文献[1]方法,但高于所提方法 2.58 ms。

相比之下,所提方法具有更高的计算效率,同时也代表所提方法具有更高的告警效率,提高了系统的工作效率。

在此基础上,对上述实验结果进行实际验证。通过监控系统对机器人进行实时监测,收集机器人传感器数据和执行状态信息。利用 Hadoop 平台进行数据处理和分析,检测异常情况并触发相应的告警。在上述设定的 IDC 安防巡检机器人的巡检环境下,基于 Hadoop 平台构建异常检测模型,通过监控系统实时采集机器人的传感器数据和执行状态信息,记录机器人的位置坐标。当机器人位置偏离预定区域,系统及时发送告警信息给相关人员。检测异常情况结果统计如表 2 所示。

表 2　IDC 安防巡检机器人网络监控系统告警结果

位置偏离预定区域次数/次	告警信息发送时延/ms
1	2.31
2	3.34
3	2.58
4	2.96
5	2.17
6	2.68
7	3.05
8	3.81

由表 2 可知,当检测到机器人位置偏离预定区域,IDC 安防巡检机器人网络监控系统告警信息发送时延最长为 3.34 ms,表明所设计系统能够及时触发相应的告警机制。运维人员接收到告警信息后进行验证和处理,还可以前往现场检查机器人位置或传感器状态,或与机器人操作人员联系确认状况。通过上述实际验证,可以确保 IDC 安防巡检机器人的网络监控系统的可靠性、及时性和准确性,提高 IDC 安全和机房设备维护的效率和可靠性。

4　结语

为了降低 IDC 运维成本,保障运维环境的安全性,提出基于 Hadoop 的 IDC 安防巡检机器人网络监控系统告警方法。首先设计低成本、高效率的 Hadoop 数据处理平台,在该平台上引入 IDC 安防巡检机器人网络监控系统,用于数据中心运维环境的全面巡检,最后以获取到的巡检监控数据为基础,将随机森林融入到建立的基于能量特征变换的无阈值告警模型中,实现 IDC 安防巡检机器人网络监控系统告警。本方法能够有效提升 IDC 安防巡检机器人网络监控系统告警的准确率及稳定度,且方法耗时较短,具有重要的实际应用意义。

参考文献:

[1] 赵庆兵,魏士源,翟小飞,等. 基于参数自回归算法的核电厂关键设备早期预警方法研究[J]. 核动力工程,2021,42(6):209-214.

[2] 林凌云,陈青,金磊,等. 基于知识图谱的变电站告警信息故障知识表示研究与应用[J]. 电力系统保护与控制,2022,50(12):90-99.

[3] 朱嘉斌. 基于 Hadoop+MPP 架构的城市轨道交通大数据中心建设方案[J]. 城市轨道交通研究,2022,25(5):54-57.

[4] 乔嘉林,黄向东,杨义繁,等. 基于着色 Petri 网的 HDFS 数据一致性建模与分析[J]. 软件学报,2021,32(10):2993-3013.

[5] 贺晓峰,廖志伟,肖华明,等. 智能巡检机器人传感与控制系统研究与设计[J]. 中国煤炭,2022,48(增刊 1):1-5.

[6] 王春露,田瑞冬,赵旭,等. ARM 处理器分支预测漏洞分析测评及新漏洞发现[J]. 西安交通大学学报,2021,55(7):71-78.

[7] 熊中敏,郭怀宇,吴月欣. 缺失数据处理方法研究综述[J]. 计算机工程与应用,2021,57(14):27-38.

[8] 吴忠强,曹碧莲,侯林成,等. 基于小波包变换和随机森林算法的光伏系统故障分类[J]. 计量学报,2021,42(12):1650-1657.

[9] 吴泽枫,李成刚,宋勇,等. 基于 NB-IoT 模块的机器人监控系统移动应用开发[J]. 机械制造与自动化,2021,50(1):161-163,197.

[10] 郑伶俊. 变电站巡检机器人系统设计与实现[J]. 机械制造与自动化,2023,52(2):162-165,188.

收稿日期:2023-08-08