



基于ECA-TCN的数据中心磁盘故障预测

张铭泉, 王宝兴

引用本文:

张铭泉, 王宝兴. 基于ECA-TCN的数据中心磁盘故障预测[J]. 智能系统学报, 2025, 20(2): 389-399.

ZHANG Mingquan, WANG Baoxing. Disk failure prediction in data centers based on ECA-TCN[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(2): 389-399.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202310043>

您可能感兴趣的其他文章

一种卷积神经网络集成的多样性度量方法

Diversity measuring method of a convolutional neural network ensemble

智能系统学报. 2021, 16(6): 1030-1038 <https://dx.doi.org/10.11992/tis.202011023>

结合卷积特征提取和路径语义的知识推理

Knowledge-based inference on convolutional feature extraction and path semantics

智能系统学报. 2021, 16(4): 729-738 <https://dx.doi.org/10.11992/tis.202008007>

新冠肺炎疫情趋势预测模型

Epidemics trend prediction model of COVID-19

智能系统学报. 2021, 16(3): 528-536 <https://dx.doi.org/10.11992/tis.202008037>

基于小样本学习的LCD产品缺陷自动检测方法

An automatic small sample learning-based detection method for LCD product defects

智能系统学报. 2020, 15(3): 560-567 <https://dx.doi.org/10.11992/tis.201904020>

BP神经网络和支持向量机相结合的电容器介损角辨识

Capacitor dielectric loss angle identification based on a BP neural network and SVM

智能系统学报. 2019, 14(1): 134-140 <https://dx.doi.org/10.11992/tis.201805034>

基于深度学习的视频预测研究综述

Review of deep learning-based video prediction

智能系统学报. 2018, 13(1): 85-96 <https://dx.doi.org/10.11992/tis.201707032>

DOI: 10.11992/tis.202310043

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20241223.1108.002>

基于 ECA-TCN 的数据中心磁盘故障预测

张铭泉^{1,2}, 王宝兴^{1,2}

(1. 华北电力大学 控制与计算机工程学院, 河北 保定 071003; 2. 华北电力大学 复杂能源系统智能计算教育部工程研究中心, 河北 保定 071003)

摘要: 随着数据中心规模的不断扩大, 磁盘故障对数据中心的运行稳定性产生越来越大的影响。当前预测方法在面对大规模、高维度和长序列的磁盘运行数据时仍存在不足。本文提出了一种高效通道注意力时间卷积网络 (efficient channel attention-temporal convolutional network, ECA-TCN) 模型, 通过结合传统卷积神经网络一维卷积的优势, 融入扩张卷积和残差结构, 并引入注意力机制, 该模型能够提高磁盘故障预测的准确性和稳定性。在实验中, 将 ECA-TCN 模型与其他经典深度学习方法进行了比较, 实验结果表明, ECA-TCN 模型在磁盘故障预测任务上具有较高的准确性和稳定性。

关键词: 磁盘故障预测; 长短时记忆网络; 循环神经网络; 扩张卷积; 高效通道注意力机制; 神经网络模型; 时间序列预测; 深度学习优化

中图分类号: TP391.4 文献标志码: A 文章编号: 1673-4785(2025)02-0389-11

中文引用格式: 张铭泉, 王宝兴. 基于 ECA-TCN 的数据中心磁盘故障预测 [J]. 智能系统学报, 2025, 20(2): 389-399.

英文引用格式: ZHANG Mingquan, WANG Baoxing. Disk failure prediction in data centers based on ECA-TCN[J]. CAAI transactions on intelligent systems, 2025, 20(2): 389-399.

Disk failure prediction in data centers based on ECA-TCN

ZHANG Mingquan^{1,2}, WANG Baoxing^{1,2}

(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China; 2. Engineering Research Center of Intelligent Computing for Complex Energy Systems Ministry of Education, North China Electric Power University, Baoding 071003, China)

Abstract: With the continuous expansion of the scale of the data center, disk failure has an increasing impact on the stability of the data center. Current prediction methods still have shortcomings in the face of large-scale, high-dimensional and long sequence of disk running data. This paper proposes an efficient channel attention-temporal convolutional network (ECA-TCN) model. By combining the advantages of one-dimensional convolution of traditional convolutional neural network, integrating dilated convolution and residual structure, and introducing attention mechanism, the model can improve the accuracy and stability of disk failure prediction. In the experiment, the ECA-TCN model is compared with other classical deep learning methods. The experimental results show that the ECA-TCN model has high accuracy and stability in the disk failure prediction task.

Keywords: disk failure prediction; long short-term memory network; recurrent neural network; dilated convolution; efficient channel attention mechanism; neural network model; time series prediction; deep learning optimization

在这个数字化世界的快速发展中, 数据中心已成为信息技术基础设施的核心。它的稳定性和

可靠性直接影响到各类数字化服务的运行。然而, 随着数据中心规模的不断扩大, 磁盘故障的影响越来越突出, 预测磁盘故障成为提升数据中心可靠性和优化维护成本的重要任务。

在硬盘故障预测领域, 已经进行了大量深入

收稿日期: 2023-10-31. 网络出版日期: 2024-12-23.

基金项目: 中央高校基本科研业务费专项项目 (2020MS122).

通信作者: 张铭泉. E-mail: mqzhang@ncepu.edu.cn.

©《智能系统学报》编辑部版权所有

的研究和实验验证。这些工作主要依赖于机器学习、深度学习和在线学习等多种算法和方法。其中,一些研究主要侧重于利用传统的机器学习方法(例如决策树^[1]、贝叶斯网络^[2-3]、支持向量机^[4-5]、回归树^[6-7]、随机森林^[8-11])来分析硬盘的自我监测、分析及报告(self-monitoring, analysis, and reporting technology, SMART)属性,进而预测硬盘的故障。另一些研究则利用深度学习技术,例如长短时记忆网络(long short-term memory, LSTM)^[12-18]、循环神经网络(recurrent neural network, RNN)^[19]、门控循环单元(gated recurrent unit, GRU)^[20]和卷积Transformer模型^[21-23]来分析时间序列数据,以捕捉并预测潜在的硬盘故障。此外,还有一些研究侧重于在线学习^[24]和迁移学习^[25-27]的方法,这些方法能够在数据的分布不断变化的情况下动态更新模型,并在不同的硬盘型号和制造商之间迁移学习模型,从而提高硬盘故障预测的准确性和可靠性。除了上述的研究方法,还有一些研究者致力于扩充数据集和增加更丰富的特征维度^[28-30]以提高硬盘故障预测的准确性。

但是,这些方法在面对大规模、高维度和长序列的数据时,往往难以达到预期的预测精度和稳定性。面对这一挑战,需要寻求一种更先进、更有效的技术手段。

为了应对这一问题,本文提出了一种基于高效通道注意力时间卷积网络(efficient channel attention-temporal convolutional network, ECA-TCN)的新型数据中心磁盘故障预测方法,并开展了以下的研究工作。

1) 磁盘故障信息的数据预处理与特征提取。深入分析了 Backblaze 公司在过去3年内公开发布的磁盘数据集。为确保数据的准确性和完整性,首先对数据进行了系统的预处理,包括数据清洗、筛选和剔除损坏磁盘等关键步骤。接下来,经过深入剖析,确定了与磁盘故障紧密相关的18个关键属性特征,并基于这些属性,利用 ECA-TCN 模型对潜在的磁盘故障进行了前瞻性的预测分析。

2) 结合卷积神经网络(convolution neural network, CNN)、扩张卷积和残差结构,为磁盘故障预测提供了一种高效的时序数据分析模型。ECA-TCN 结合了传统 CNN 的优势并融入了扩张卷积和残差结构,这使其能够深入挖掘长期的时序负荷数据中的非线性跨时依赖关系。残差结构的引入,不仅提升了多层神经网络的反馈机制,还加速了模型的收敛过程。相较于其他如 RNN 等传

统序列模型, ECA-TCN 在磁盘故障预测方面表现出了一系列优势,例如:并行计算的能力、灵活调节的感受野大小、稳健的梯度变化、较低的内存需求以及处理各种长度输入数据的适应性。

3) 在时序数据处理上, ECA-TCN 通过引入高效通道注意力(efficient channel attention, ECA)模块显著增强了其注意力机制。ECA 模块是一种高效的通道注意力模块,它通过一维卷积实现局部跨信道交互,具有较少的参数和明显的效果增益。在时间卷积网络(temporal convolutional network, TCN)中添加 ECA 模块可以提高模型的性能和效率。

这种方法结合了深度学习在时序数据分析上的强大能力,以及 ECA-TCN 模型在捕捉动态特征和跨通道关联性方面的优势,使得预测结果在精确率、召回率和 F1 分数等多项评价指标上均显著优于现有方法。

1 相关工作

1.1 基于机器学习的磁盘故障预测

Rincón 等^[1]研究了在异构环境中利用磁盘的 SMART 属性进行磁盘故障预测,使用机器学习技术构建了一个基于决策树的预测模型。该方法能够预测异构磁盘群体中大约 52% 的磁盘故障,具有较高的准确性和较低的虚报率。Chaves 等^[2]和 Hamerly 等^[3]研究了基于贝叶斯网络的硬盘故障预测方法,实验结果显示,与基线模型相比,提出的方法在平均数和中位数的均方误差上分别提升了 28.3% 和 17.6%。Zhang 等^[4]研究了基于混集成学习的硬盘故障预测,将多种机器学习方法相结合,包括逻辑回归、K 最近邻、支持向量机、朴素贝叶斯、随机森林,以提高硬盘故障预测的准确性和可靠性。Fulp 等^[5]提出使用一个滑动窗口的子序列作为支持向量机(support vector machine, SVM)的输入来预测磁盘故障。Li 等^[6]首先基于分类树和回归树的硬盘故障预测模型,之后使用决策树和梯度提升回归树^[7]来预测硬盘故障,并通过实验验证了这些模型在预测性能、稳定性和可解释性方面的优势,同时具有较低的误报率。Shen 等^[8]基于部分投票随机森林(random fores, RF)的方法针对硬盘驱动器的多样性和不同故障类型之间的差异,通过对健康样本进行粗略分类,提高了故障预测的准确性。李国^[9]提出一种基于变权重随机森林模型的优化方法,采用计算特征属性值和硬盘故障的相关性对原始数据集进行降维处理,为解决预测硬盘故障的问题提

供了一种借鉴思路。乔旭坤等^[10]在统一的实验平台中对多种机器学习算法进行了预测对比研究,实验结果表明随机森林模型和梯度提升决策树模型不仅预测精度很高而且对不同规模的样本集具有很强的泛化性。李新鹏等^[11]提出了一种基于自适应加权梯度提升决策树(bagging-gradient boosting decision tree, Bagging-GBDT)算法的磁盘故障预测模型,采用基于聚类的分层欠采样方法进行健康磁盘样本抽样,并结合多个GBDT子分类模型通过加权硬投票进行预测,实验结果显示其在不平衡数据集上的召回率和故障预测准确性都得到了显著提升。

1.2 基于深度学习的磁盘故障预测

Shi等^[12]提出了一种改进的机械硬盘故障预警算法,通过特征选择、生成对抗网络和卷积LSTM模型对硬盘故障进行准确检测和预警。在机械硬盘故障预警任务中取得了优异的精确度、召回率和AUC(area under curve)值等性能指标。结果表明,在移动边缘计算环境中,使用LSTM循环神经网络的方法可以提高硬盘驱动器(hard disk drive, HDD)的可靠性^[13]。Yang等^[14]和兰云龙等^[15]采用LSTM算法进行磁盘故障预测,通过训练模型对磁盘状态进行分类,同时进行正负样本平衡和特征数据处理。姜少彬等^[16]为解决传统方法在不均衡数据集的过拟合问题,该研究提出了一种结合LSTM和生成式对抗网络的非监督对抗学习方法进行硬盘故障预测,实验证明其在异常检测的召回率、精确度和泛化能力上均优于传统分类器。Hu等^[17]引入时间序列分析的概念,通过使用基于LSTM的神经网络模型考虑磁盘的连续运行记录,平均预测精度为86.31%,准确预测磁盘在未来15 d内的故障。Wu等^[18]提出了基于多通道卷积神经网络和LSTM(multi-scale convolutional neural network-long short-term memory network, MCCNN-LSTM)模型,能够在长期预测范围和长时间窗口中实现更准确地预测。Xu等^[19]用RNN来分析SMART属性的时间序列数据,建立模型进行硬盘健康度评估和故障预测。通过利用循环连接在隐藏层之间传播时间序列信号,保留了SMART属性的历史信息。Hai等^[20]提出了一种基于GRU神经网络和时间序列生成对抗网络(time-series generative adversarial network, TimeGAN)对抗网络的硬盘故障预测方法,GRU神经网络可以适应长硬盘数据序列的影响,而TimeGAN可以解决数据不平衡问题,具有很好的动机和合理性。Xu等^[21]提出了一种面向大规模网络存储系

统磁盘故障预测的卷积Transformer模型ConvTrans-TPS(convolutional Transformer model for disk failure prediction in large-scale network storage systems)模型与近年来流行的卷积神经网络-长短期记忆网络(convolutional neural network-long short-term memory, CNN-LSTM)模型相比,所提方法的F1分数和马修相关系数值分别提升了4%和5%,提高了预测精度。Sun等^[22]提出了一种新颖的基于深度学习的硬件故障预测方案。其中包括属性分布归一化、基于时间卷积神经网络的模型和改进的损失函数,并取得了良好的性能表现。Guan等^[23]提出一种名为马尔可夫决策过程-模型无关元学习(Markov decision process-model agnostic meta learning, MDP-MAML)的磁盘故障预测模型。MDP解决了数据不均匀划分和按时间划分的挑战,而MAML解决了针对多个问题小样本学习的问题。

1.3 基于在线学习和迁移学习的磁盘故障预测

Xiao等^[24]提出了一种基于在线学习的磁盘故障预测模型,即在线随机森林(online random forest, ORF)算法。ORF算法支持数据的顺序到达,动态适应智能数据的分布,克服了“模型老化”问题,在线学习方法有效地解决了在在线环境下训练磁盘故障预测模型的挑战。Zhang等^[25]提出了一种名为基于迁移学习的少数磁盘故障预测(transfer learning based failure prediction for minority disks, TLDFP)的算法,它包括迁移学习算法TrAdaBoost(boosting for transfer learning)和实例到磁盘的映射算法。TLDFP通过减少源域和目标域之间的数据分布差异,建立不同磁盘模型之间的关系,并提高少数磁盘故障预测的性能。TLDFP能够同时实现高的故障检测率和低的误报率,且在AUC-ROC(area under the receiver operating characteristic curve)曲线方面也有较好的分类能力。Han等^[26]设计了磁盘故障预测的通用流挖掘框架,基于3种关键技术,即在线标记、概念漂移感知训练和通用预测,其主要目标是支持各种机器学习算法,实现了少数磁盘模型的在线迁移学习,并实现了相当高的流处理性能。关东海等^[27]提出了一种结合迁移学习和主动学习的硬盘故障预测方法,通过处理SMART属性数据、主成分分析确定健康属性、归一化及堆叠转换,然后采用迁移算法生成模型来预测目标域硬盘的故障,从而提高预测准确性。

1.4 基于扩充数据集的方法

Lu等^[28]则证明了在SMART属性的基础上

加入 I/O Performance 属性和 Location 属性可以提高预测准确率,证明了引入更高维度的特征信息有助于更全面地找出故障磁盘。Han^[29] 等进一步考虑了固态硬盘集群存在的关联故障问题,将磁盘故障预测问题与实际数据中心相结合,并根据故障清单给出了机房配置的指导建议。Wang 等^[30] 设计了一种数据增强方法,该方法可以有效地生成高质量的故障样本,它使用基于置信度的方法过滤掉低质量的生成样本,缓解训练集的数据不平衡问题,从而有效提高任何有监督故障检测模型的性能。

1.5 现有工作的缺点

在当前研究中存在一些不足,需进一步完善。虽然机器学习、深度学习和在线学习在磁盘故障预测上已有所成果,但还面临挑战。数据不平衡问题导致预测模型在正常和故障磁盘的识别上存在偏差。从高维稀疏 SMART 属性数据中提取有价值信息也具有难度。某些深度学习方法虽准确,但模型复杂,消耗大量计算资源和训练时间。在线学习需适应数据流变化,迁移学习则要确保源与目标领域数据关系的持续性。模型更新机制的进一步研究也很关键,以应对磁盘环境和使用的变化。而且,现有方法在实时预测上需提高计算效率。

与此相比,本文提出的 ECA-TCN 模型作出了改进。该模型通过生成基于邻近样本的合成样本来平衡数据集,这样可降低数据不平衡对预测模型的影响,并增强对故障硬盘的预测性能。引入的注意力模块,在一定程度上解决了高维稀疏数据的挑战。与现有方法相比,ECA-TCN 考虑了多个 SMART 属性的时间序列信息,通过残差模块构建了更深的网络结构,从而更好地捕捉非线性特征和长期依赖。更为重要的是,ECA-TCN 具有并行计算的优点,提高了计算效率。ECA-TCN 既有强大的表达力也具备适应性,满足实时预测需求。

2 数据处理

2.1 数据预处理

本文采用近 3 年 Backblaze 数据中心磁盘 SMART 信息数据集,Backblaze 数据集详细记录了每天数据中心磁盘的运行状态和性能指标,主要以序列化数据格式存储。该数据集包括了丰富的文本信息,如磁盘型号、运行时间以及各种 SMART 属性值等。本文选用近 3 年平均故障率最高的磁盘型号 ST4000DM000 来进行磁盘故障预测分析,具体信息见表 1。

表 1 Backblaze 磁盘数据集运行故障统计
Table 1 Backblaze disk dataset run failure statistics

年份	磁盘总数	平均运行时间/d	故障磁盘数	故障率/%
2020	18 939	1 870	269	1.41
2021	18 611	2 231	339	1.80
2022	18 246	2 577	633	3.45

数据预处理流程主要包括丢弃不合格磁盘、特征值选择、数据归一化、基于滑动窗口的时序数据集生成等,数据预处理流程如图 1 所示。

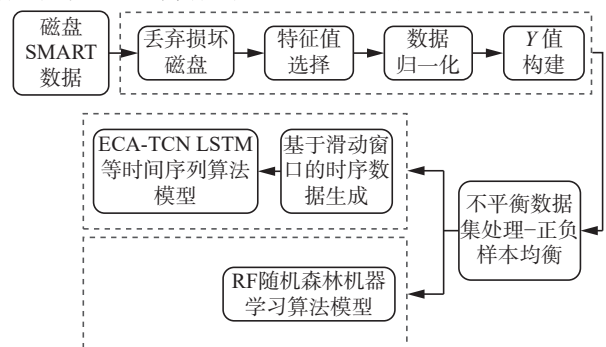


图 1 数据预处理流程

Fig. 1 Data preprocessing flow

2.1.1 丢弃不合格磁盘

由于数据集中存在大量的缺失值,且部分磁盘的运行天数较短,无法构建用于磁盘故障预测的时序数据集,因此筛选并移除质量不佳的硬盘是非常必要的。本文首先删除空值大于 40% 的属性列。其次,基于硬盘的启动天数和指定时间窗口内的缺失值数量,识别并排除启动天数不足 90 或缺失窗口达到 30 d 的磁盘数据。

2.1.2 特征值选择

磁盘 SMART 信息包含最多 40 个数据点。这些数据能最直接反映硬盘当前的工作状况。在 SMART 属性数据中,记录了诸如工作时长、通电次数、盘片温度、坏块数、寿命余量、PE 数、链路错误次数、写入量、读出量等重要运行信息。然而,部分属性与磁盘故障无关,因为它们固定的或变化不明显。因此,本文通过相关性检验,选择与磁盘故障最相关的属性进行预测研究。表 2 是选择的属性列表。选定的 SMART 属性涵盖了硬盘的多方面性能和健康状态。例如,读取错误率(表 2 中编号: 001)和寻道错误率(007)反映了数据读取和磁头寻道的准确性,而重定位磁区计数(005)和等候重定的扇区计数(197)则揭示了硬盘磁区的健康状况。硬盘加电时间(009)记录了硬盘的运行时间,而设备开关计数(012)和电机起停次计数(004)则反映了硬盘的使用周期。SATA(serial advanced technology attachment)

降级运行计数 (183)、UltraDMA(ultra direct memory access) 通讯循环冗余校验 (cyclic redundancy check, CRC) 错误 (199) 和通信超时 (188) 展示了数据传输和通信的稳定性。终端校验出错 (184)、报告不可纠正错误 (187) 和无法校正的扇区计数 (198) 提供了硬盘错误的重要指标。最后, 电源关闭磁头收回计数 (192) 和气流温度 (190) 提供了硬盘的物理状态信息。这些属性综合展现了硬盘的健康和性能状态, 有助于算法及时识别和预防硬盘潜在问题。

表2 Backblaze 磁盘数据集运行故障统计
Table 2 Backblaze disk dataset run failure statistics

编号	SMART属性名称	编号	SMART属性名称
001	smart_1_raw	012	smart_12_raw
004	smart_4_raw	183	smart_183_raw
005	smart_5_raw	197	smart_197_raw
009	smart_9_raw	199	smart_199_raw
184	smart_184_raw	007	smart_7_raw
187	smart_187_raw	198	smart_198_raw
240	smart_240_raw	188	smart_188_raw
241	smart_241_raw	192	smart_192_raw
242	smart_242_raw	190	smart_190_raw

2.1.3 数据归一化

磁盘不同属性具有不同输出范围, 尺度大的特征可能主导模型训练, 降低性能。本文采用最大最小归一化 (MinMaxScaler) 来使数据特征在相同数量级, 以便深度学习任务更好地发挥作用。对磁盘 SMART 属性进行规范化:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

式中: X_{norm} 为归一化后得到的数值, X 为待归一的数据, X_{\min} 表示数据中最小值, X_{\max} 表示数据中最大值。对数据进行了按列归一化, 将不同的属性数据归一化到 $[0, 1]$ 区间, 本文也尝试了其他归一化方法例如 Z-score 标准化, 在本文算法模型下, 最大最小归一化取得了更好的预测结果。

2.2 基于滑动窗口的时序数据生成

在处理时序数据时, 一个常用的策略是使用滑动窗口来生成数据。首先使用 MinMaxScaler 对数据进行标准化。接下来, 开始实现滑动窗口的功能。对于每一行, 通过滑动窗口截取的方式, 生成过去的时序信息。例如, 对于一个 T d 的窗口, 当处理第 N 天的数据时, 生成的时序数据将包含 $(N-T, N]$ 的数据。滑动窗口如图2所示。

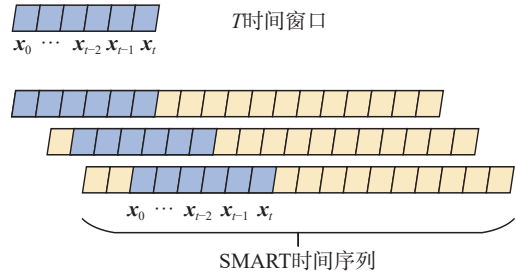


图2 滑动窗口

Fig. 2 Sliding window

由于在创建滑动窗口时, 数据的前部分可能会因为缺少足够的过去信息而产生缺失值, 因此需要删除所有包含缺失值的行, 并对数据框进行排序, 以确保后续处理过程中正确使用数据的顺序。滑动窗口方法成功地为模型提供了丰富的历史信息, 以帮助模型更好地理解时序数据的内在规律。

2.3 Y值构建

本文构建一个与输入数据相对应的目标变量 y_{train} 。 y_{train} 的每个值代表其对应的磁盘在接下来的 7 d 内是否会发生故障。如果在某磁盘的观察时间点后的 7 d 内发生了故障, 那么该磁盘在该观察时间点的 y_{train} 值为 1; 反之, 如果在接下来的 7 d 内没有发生故障, 其 y_{train} 值为 0。

通过这种 Y 值的构建方式, 模型可以更加专注于预测短期内磁盘的故障风险, 从而为及时的维护和故障预防提供有力的支持。

2.4 不平衡数据集的处理

在磁盘故障预测应用中, 样本数据类别极度不平衡。故障磁盘的数量远小于正常运行的磁盘, 在本文样本数据中故障磁盘只占到所有样本磁盘的 2.23%, 且生成滑动窗口数据后, 由于故障磁盘在大部分时间内是正常运行的, 样本差距将被进一步放大, 这种数据偏斜可能导致模型在预测时产生偏见, 特别是对于那些少数的类别。因此, 对于不平衡数据集的处理变得尤为重要。

本文定义了一个采样策略: 将少数类的数量增加到多数类的 10%。为此, 采用了 2 种技术:

过采样 (over-sampling) 使用 SMOTE (synthetic minority over-sampling technique) 方法对少数类别进行过采样。SMOTE 是一种流行的过采样方法, 对于少数类样本, 随机选择 1 个样本, 然后从其 k 个近邻中选择 1 个近邻。之后, 在这 2 个样本之间的连线上随机选择 1 个点作为合成的新样本。

$$\mathbf{x}_{\text{new}} = \mathbf{x}_i + \lambda \times (\mathbf{x}_{\text{nn}} - \mathbf{x}_i)$$

式中: \mathbf{x}_{new} 为合成的新样本, \mathbf{x}_i 为从少数类中随机

选择的一个样本, x_m 为从 x_i 的 k 个近邻中随机选择的一个近邻样本, λ 为一个介于 0 和 1 之间的随机数。

欠采样 (under-sampling) 使用随机欠采样方法对多数类进行采样, 从多数类中随机删除一些样本, 从而使得多数类的样本数量降低, 确保多数类样本的数量是少数类的 10 倍。

$$n_{\text{new}} = \frac{n_{\text{minority}}}{r}$$

式中: n_{new} 为欠采样后多数类的目标样本数量, n_{minority} 为少数类的样本数量, r 为所希望的多数类和少数类之间的比例。

综上所述, 通过组合过采样和欠采样的策略, 成功地处理了不平衡数据集, 为后续模型训练创造了更有利的条件。

3 基于 ECA-TCN 的磁盘故障预测模型

基于 ECA-TCN 的磁盘故障预测模型如图 3 所示, 模型分为 3 部分。输入层负责将预处理后的硬盘故障 SMART 特征向量转换为主网络的输入数据, 随后主干网络从这些具有多样性的输入序列中提炼出非线性特性, 最终输出层将主网络生成的多维特性序列转化为表示是否出现故障的标签输出。

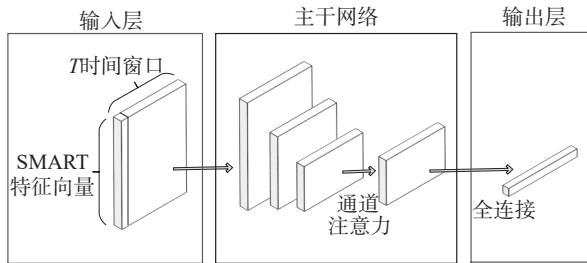


图 3 ECA-TCN 磁盘故障预测模型

Fig. 3 ECA-TCN disk failure prediction model

3.1 TCN 模块构建

TCN 是一种用于处理序列数据的时序卷积网络。TCN 的主要特点是:

1) 使用一维卷积层来提取序列的局部和全局特征, 而不是使用 RNN 或 LSTM。

2) 使用因果卷积 (causal convolution) 来保证输出只依赖于当前和过去的输入, 而不是未来的输入, 从而避免信息泄露。

3) 使用扩张卷积 (dilated convolution) 来增加感受野 (receptive field), 即每个输出能够覆盖的输入序列的长度, 从而捕捉更长期的依赖关系。扩张卷积网络结构如图 4 所示, 选取历史时刻的输入 x_0, x_1, \dots, x_{t-1} 作为模型 t 时刻的输入 x_t , 满足时序预测的输入条件, 每一个输入为磁盘某一个

时间点的 SMART 属性向量, 输出为 $\hat{y}_0, \hat{y}_1, \dots, \hat{y}_t$ 对应磁盘的故障标签, d 为扩张系数。本研究采用扩张卷积, 扩张系数代表了在卷积核中的间隔大小, 允许在不增加计算成本的情况下处理更长的时间序列依赖, 从而提高预测精度。

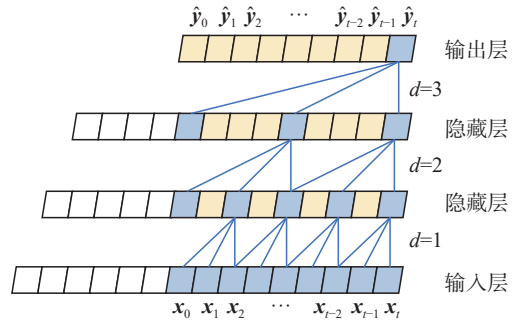


图 4 扩张卷积网络结构

Fig. 4 Dilated convolutional network architecture

4) 使用残差连接 (residual connection) 来增加网络深度, 从而提高模型的表达能力和泛化能力。如图 5 所示, 残差结构允许神经网络层级之间的梯度直接传播。这种梯度传播机制减轻了训练过程中的梯度消失问题, 并且有助于更快地收敛和获得更好的性能。

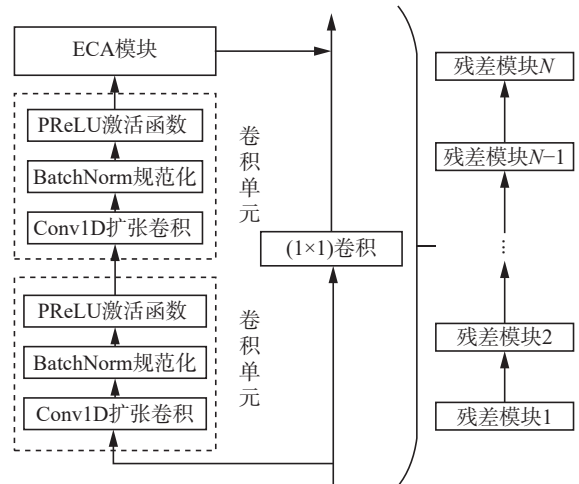


图 5 残差单元结构

Fig. 5 Residual cell structure

3.2 ECA 网络模块构建

ECA-Net (efficient channel attention network) 提出了一种轻量级的自适应通道注意力机制^[31], 该机制可以为卷积神经网络中的特征通道赋予权重, 从而提高了卷积网络的性能^[32], 它的设计原则是在保持通道注意力的有效性的同时, 减少参数和计算成本。其核心思想是通过一维卷积来捕获不同通道之间的交互。

在现代深度学习模型中, 如何有效地提取和利用特征间的依赖关系是一个关键的挑战。尤其在磁盘 SMART 信息时间序列数据中, 这种依赖

性可能存在于长时间跨度之间, 因此需要特殊的结构来捕捉。为了解决这个问题, 本文使用了 TCN, 通过扩张卷积来提高其在捕捉长期依赖关系方面的能力。

然而, 除了时间维度上的依赖关系, 磁盘 SMART 信息特征通道之间也存在潜在的依赖关系, 这通常被忽视。为了充分挖掘这些依赖关系并提高模型性能, 提出了将 ECA-Net 与 TCN 相结合的方法。ECA-Net, 作为一种有效的通道注意力机制, 可以为每个特征通道分配一个自适应权重。

注意力值在计算过程中, 自注意力层的打分函数采用缩放点积的方式, 其输出为

$$A_{\text{attention}}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}$$

式中: \mathbf{K} 和 \mathbf{V} 代表键值对 (key, value); \mathbf{Q} 表示目标数据; d_k 表示 \mathbf{Q} 的维度。要得到 $A_{\text{attention}}$ 的计算值需要满足 $\mathbf{Q} = \mathbf{K} = \mathbf{V}$ 。ECA-TCN 的注意力机制网络结构如图 6 所示。TCN 的输出为 $\hat{\mathbf{y}}_i$, 所对应的特征向量, 作为注意力机制层的输入 $\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_n$, 输入具有多种不同的特征, 并且这些特征与时间序列有着显著的关联。在预测结果中, 每个向量都占有不同的权重。因此, 注意力机制的作用在于强调在传递故障预测数据信息中权重较高的特征属性。这是通过计算向量间的相关性权重并进行加权求和, 从而得到注意力值来实现的。注意力机制能够自动识别并关注那些对硬盘故障预测更为重要的 SMART 属性, 从而在一定程度上解决了高维稀疏数据带来的挑战。

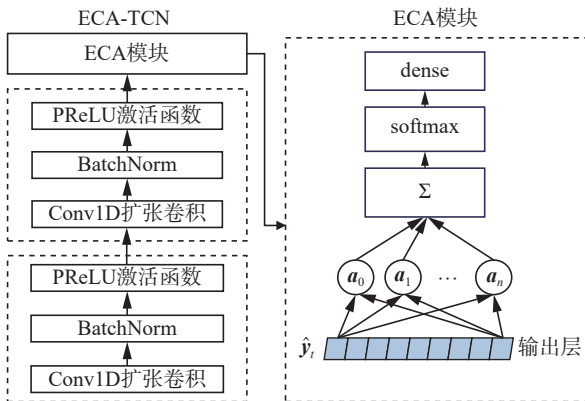


图 6 ECA 模块
Fig. 6 ECA module

结合 ECA-Net 和 TCN 的优势在于: TCN 可以有效地捕捉时间维度上的长期依赖, 而 ECA-Net 则能够增强模型对于不同特征通道的区分能力。这种结合不仅增强了模型的表示能力, 同时保持了模型的计算效率。实验结果表明, 本文设

计的 ECA-TCN 模型在磁盘故障预测任务上取得了显著的性能提升。这证明了通道注意力机制和时序建模的结合能够为深度学习模型带来额外的好处。

3.3 评价指标

在评估 ECA-TCN 模型在磁盘故障预测上的性能时, 本文采用了 4 种核心的评价指标: 精确率、召回率、F1 分数和马修斯相关系数 (Matthews correlation coefficient, MCC)。这些指标提供了模型效果的全面视图, 并有助于更好地理解模型在各种磁盘预测场景情境下的性能。

精确率评估模型预测为正例的样本中实际为正例的比例:

$$I_{\text{Precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}$$

式中: N_{TP} 是真正例数, N_{FP} 是假正例数。精确率表示当模型预测某个磁盘会出现故障时, 这个预测正确的百分比是多少。

召回率或灵敏度评估实际为正例的样本中被模型预测为正例的比例:

$$I_{\text{Recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}$$

式中 N_{FN} 是假负例数。对于磁盘故障预测, 召回率描述的是所有真实出现故障的磁盘中, 有多少百分比被模型正确预测出来了。

F1 分数是精确率和召回率的调和平均, 能够平衡两者之间的关系, 特别适用于样本不平衡的情况。

$$F1_{\text{Score}} = 2 \times \frac{I_{\text{Precision}} \times I_{\text{Recall}}}{I_{\text{Precision}} + I_{\text{Recall}}}$$

在一个系统中, 磁盘故障是少数事件, 那么单一地看精确率或召回率可能会导致偏见。F1 分数能够确保在这种情境下, 模型的性能评估是公平和平衡的。

马修系数是一个在二元分类中评估模型的指标, 它考虑了真正例、假正例、真负例和假负例。

$$M_{\text{cc}} = \frac{(N_{\text{TP}} \times N_{\text{TN}}) - (N_{\text{FP}} \times N_{\text{FN}})}{\sqrt{(N_{\text{TP}} + N_{\text{FP}})(N_{\text{TP}} + N_{\text{FN}})(N_{\text{TN}} + N_{\text{FP}})(N_{\text{TN}} + N_{\text{FN}})}}$$

M_{cc} 提供了一个在 $[-1, 1]$ 范围内的评分, 其中 $M_{\text{cc}} = 1$ 代表完美的预测, $M_{\text{cc}} = 0$ 表示随机预测, 而 $M_{\text{cc}} = -1$ 表示完全相反的预测。在磁盘故障预测中, M_{cc} 可以提供一个全面的视图, 关于模型在正例和负例上的整体表现。例如, M_{cc} 接近 1 表示模型在预测磁盘故障和非故障上都表现得很好。

4 实验结果及其对比

4.1 实验平台介绍

本研究的所有深度学习实验均在自组装的 Linux 工作站上完成。工作站选用 Ubuntu 20.04.5 LTS 作为操作系统,其稳定的内核和丰富的软件库能够提供可靠的环境,满足深度学习研究的需求。处理器为 Intel(R) Xeon(R) Gold 6330 CPU 28 核心 56 线程。系统内存为 504 GB DDR4 3200 MHz。显卡为 6 张 NVIDIA RTX3090 24 GB,使用 CUDA 加速 PyTorch 深度学习框架。所有实验数据均在此 Ubuntu 环境中采集并处理。

4.2 模型训练与验证

为了深入全面评估磁盘故障预测的能力,本文采用了 4 种主流分类器:循环神经网络、长短时记忆网络、时间卷积网络以及集成了高效通道注意力模块的时间卷积网络。此外,研究还纳入经典机器学习 Bagging-GBDT 算法随机森林,以进行比较分析。

在循环神经网络的实验设置中,选用单一隐藏层,隐藏状态大小设为 64。学习率定为 0.001,批量大小为 256,进行了 300 个训练周期。

长短时记忆网络的配置保持学习率 0.001、批量大小 256 和 300 训练周期不变,同时为降低过拟合风险,在网络结构中引入了 0.1 的 Dropout 比例。

时间卷积网络及其高效通道注意力版本的训练参数保持一致,包括批量大小 256、学习率 0.001 和 300 个训练周期。

对于随机森林算法,根据其特点调整了相应的参数配置,以确保能够在磁盘故障预测中达到最佳效果。

本文在训练和验证阶段均采用精确率、召回率、F1 分数和马修斯相关系数等评价指标,以确保对所有模型进行全面且深入的评估。

4.3 实验结果与对比

4.3.1 时间序列算法模型下预测结果对比分析

本小节中重点探讨磁盘故障预测实验结果和分析,实验过程中,各个模型都使用了相同的测试数据集进行预测。在此基础上,使用了经过参数调优后的 LSTM 和 RNN 模型,并与未使用 ECA 模块的 TCN 以及使用了 ECA 模块的 ECA-TCN 进行了对比分析。使用 40 d 的时间窗口,4 种算法模型在 4 种评价指标下的实验数据如表 3、图 7 所示。实验结果表明,ECA-TCN 模型与 TCN 模型在各项指标上均展现出优异的表现,

特别是在精确率和 MCC 上,其表现显著优于其他 2 种模型。这表明 TCN 模型在提高磁盘故障预测任务上的性能方面成效显著。

表 3 不同算法模型预测结果

Table 3 Prediction results of different algorithm models

模型	评价指标			
	精确率	召回率	F1分数	MCC
RNN	0.740 7	0.561 2	0.638 6	0.614 8
LSTM	0.878 1	0.711 9	0.786 3	0.772 4
TCN	0.952 9	0.680 9	0.794 2	0.795 1
RF	0.947 0	0.662 0	0.779 3	0.775 8
ECA-TCN	0.963 7	0.722 3	0.825 7	0.821 0



图 7 不同算法模型预测结果对比

Fig. 7 Comparison of prediction results of different algorithm models

与此同时,经过参数优化的 LSTM 模型也呈现出良好的表现,尤其在召回率上表现出色,但在 F1 分数、精确率以及 MCC 方面相较于 ECA-TCN 模型略显不足。这表明,在磁盘故障预测任务中,复杂的网络结构确实能够在一定程度上提升模型性能。

相较之下,RNN 在多数评价指标上的表现较为平庸。这可能是由于 RNN 在处理长序列依赖时容易受到梯度消失或梯度爆炸问题的影响,导致其在学习长序列特性时的效果不尽如人意。在磁盘故障预测任务中,长期的历史信息对于准确预测未来故障状态具有至关重要的意义。由于 LSTM 和 TCN 等模型具备较好的内部结构和门控机制,能够较好地解决这些问题,因此在性能上超越了 RNN。

尽管随机森林作为经典的机器学习算法,在许多情况下仍表现出色,但在处理具有高度复杂性和多变性的磁盘数据时,它的性能相比 ECA-TCN 有所不足。这一差距主要体现在对异常模式的识别能力和处理高维度数据的效率上。ECA-TCN 通过其独特的时间卷积网络结构,加上有效的注意力机制,能够更准确地捕捉时间序列数据中的关键特征,从而在预测未来的磁盘故障方面

展现出更高的准确性。

ECA-TCN 在处理大规模数据集方面表现出色,这得益于其引入的通道注意力机制。该机制允许模型根据不同的 SMART 属性赋予不同的权重,使其能够更专注于与硬盘故障高度相关的属性,同时忽略不相关或噪声较大的属性。这有助于解决高维稀疏数据的挑战,提高故障预测的准确性和鲁棒性。

总体而言,实验结果表明,ECA-TCN 凭借其强大的长期依赖捕捉能力和高效的注意力机制,在磁盘故障预测任务上表现出更高的准确性和可靠性。这种模型通过结合 TCN 的时序处理优势和 ECA 机制的空间特征聚焦,实现了对磁盘运行

数据中复杂模式的深入理解,显著提高了故障预测的准确度。此外,ECA-TCN 在减少误报和漏报方面也展现了卓越性能,这对于提升系统稳定性和减少维护成本至关重要。这些优点使得 ECA-TCN 成为未来磁盘故障预测和类似领域研究的一个有价值的参考点。

4.3.2 不同时间窗口下预测结果对比分析

本小节分析了不同时间窗口长度对磁盘故障预测准确性的影响。具体地,使用了不同长度的时间窗口,如 10、30、40、50、70、90 d,并分别用 RNN、LSTM、TCN 和 ECA-TCN 4 种模型对其进行了预测。通过 MCC 评价在正例和负例上的整体表现,预测结果如表 4、图 8 所示。

表 4 不同时间窗口下模型 MCC 预测结果
Table 4 Prediction results of model MCC under different time windows

模型	时间窗口/d					
	10	30	40	50	70	90
TCN	0.693 9	0.757 4	0.795 1	0.850 3	0.875 2	0.885 4
LSTM	0.616 1	0.735 4	0.772 4	0.805 7	0.820 3	0.827 3
RNN	0.542 5	0.589 7	0.614 8	0.590 9	0.555 8	0.571 7
ECA-TCN	0.727 9	0.783 2	0.821 0	0.855 3	0.891 1	0.903 7

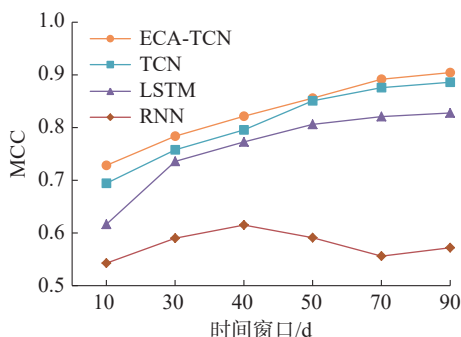


图 8 不同时间窗口下模型 MCC 预测结果对比

Fig. 8 Comparison of MCC prediction results of models under different time windows

实验结果表明,在 TCN 和 LSTM 网络模型下,时间窗口越长,磁盘故障的预测准确性确实有所提高。这是因为长时间窗口能提供更多的历史信息,从而增强模型在捕捉磁盘行为模式方面的能力。然而,随着时间窗口的进一步增长,预测准确性的增长幅度逐渐减小,显示出存在某种瓶颈效应。这意味着,尽管长时间窗口能提供更多信息,但是超过一定长度后,额外的历史数据可能不再为模型带来显著的性能提升。

值得注意的是在 RNN 模型下,随着时间窗口的增加,模型预测准确度出现了下降和波动,RNN 处理长序列时,会遇到梯度消失或梯度爆炸

的问题。这意味着,当序列很长时,RNN 可能会失去捕获序列中早期事件的能力。

ECA-TCN 通过结合 TCN 能够有效捕捉和学习硬盘 SMART 属性的时序特征和潜在的故障模式。由于 TCN 模型具有较好的计算效率,ECA-TCN 在处理大量的时序数据时仍能快速响应和高效运算。与传统的 RNN 相比,TCN 在处理长时间序列数据时能够避免梯度消失或梯度爆炸的问题,保证了模型的稳定性。同时,TCN 的因果卷积结构确保了信息只从过去流向未来,这为硬盘故障预测提供了更加准确和实时的预测结果。

综上所述,实验结果表明,在 TCN 和 LSTM 网络模型下,使用较长的时间窗口可以提高磁盘故障预测的准确性,因为这样能够提供更多的历史信息,增强模型捕捉磁盘行为模式的能力。然而,当时间窗口长度超过一定阈值后,预测准确性的提升幅度会减少,显示出瓶颈效应。特别地,在 RNN 模型中,长时间序列处理存在梯度消失或梯度爆炸的问题,导致预测准确度下降。而 ECA-TCN 模型通过有效结合时序卷积网络的优势,避免了这些问题,保证了在长时间序列数据处理中的稳定性和高效性,表现出所有模型中最优的性能。这些发现对于优化磁盘故障预测模型

和提前采取预防措施具有重要意义。

5 结束语

本文研究了基于ECA-TCN的数据中心磁盘故障预测方法,验证了其在召回率和MCC等关键性能指标上的显著优势。ECA-TCN模型能有效捕捉SMART属性间的复杂关联,并关注关键时序特征,展现出优越的性能和实用价值。尽管ECA-TCN模型展现出了优异的性能,但其应用仍受限于对特定数据集的依赖性,并且在处理极端情况和罕见故障模式时可能面临挑战。未来工作将探索ECA-TCN与其他机器学习技术的结合,并开发更灵活、适应性强的预测模型,以适应数据中心的动态变化,进一步优化故障预测的准确性和效率。

参考文献:

- [1] RINCÓN C C A, PÂRIS J F, VILALTA R, et al. Disk failure prediction in heterogeneous environments[C]//2017 International Symposium on Performance Evaluation of Computer and Telecommunication Systems. Seattle: IEEE, 2017: 1–7.
- [2] CHAVES I C, DE PAULA M R P, LEITE L G M, et al. BaNHFaP: a Bayesian network based failure prediction approach for hard disk drives[C]//2016 5th Brazilian Conference on Intelligent Systems. Recife: IEEE, 2016: 427–432.
- [3] HAMERLY G, ELKAN C. Bayesian approaches to failure prediction for disk drives[C]//Proceedings of the Eighteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 2001: 202–209.
- [4] ZHANG Mingyu, GE Wenqiang, TANG Ruichun, et al. Hard disk failure prediction based on blending ensemble learning[J]. *Applied sciences*, 2023, 13(5): 3288.
- [5] FULP E W, FINK G A, HAACK J. Predicting computer system failures using support vector machines[C]//Proceedings of the First USENIX conference on Analysis of system logs. USA: USENIX Association, 2008: 5.
- [6] LI Jing, JI Xinpui, JIA Yuhuan, et al. Hard drive failure prediction using classification and regression trees[C]//2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. Atlanta: IEEE, 2014: 383–394.
- [7] LI Jing, STONES R J, WANG Gang, et al. Hard drive failure prediction using decision trees[J]. *Reliability engineering & system safety*, 2017, 164: 55–65.
- [8] SHEN Jing, WAN Jian, LIM S J, et al. Random-forest-based failure prediction for hard disk drives[J]. *International journal of distributed sensor networks*, 2018, 14(11): 155014771880648.
- [9] 李国, 常甜甜, 李静. 基于变权重随机森林的硬盘故障预测方法[J]. *计算机工程与设计*, 2021, 42(10): 2988–2994.
- LI Guo, CHANG Tiantian, LI Jing. Hard disk failure prediction method based on variable weight random forest[J]. *Computer engineering and design*, 2021, 42(10): 2988–2994.
- [10] 乔旭坤, 李顺, 李君, 等. 基于机器学习的硬盘故障预测研究[J]. *计算机技术与发展*, 2022, 32(6): 215–220.
- QIAO Xukun, LI Shun, LI Jun, et al. Research on hard disk failure prediction based on machine learning[J]. *Computer technology and development*, 2022, 32(6): 215–220.
- [11] 李新鹏, 高欣, 何杨, 等. 不平衡数据集下基于自适应加权 Bagging-GBDT 算法的磁盘故障预测模型[J]. *微电子学与计算机*, 2020, 37(3): 14–19.
- LI Xinpeng, GAO Xin, HE Yang, et al. Prediction model of disk failure based on adaptive weighted Bagging-GBDT algorithm under imbalanced dataset[J]. *Microelectronics & computer*, 2020, 37(3): 14–19.
- [12] SHI Junjie, DU Jing, REN Yingwen, et al. Convolution-LSTM-based mechanical hard disk failure prediction by sensoring S. M. A. R. T. indicators[J]. *Journal of sensors*, 2022(5): 7832117.
- [13] SHEN Jing, REN Yongjian, WAN Jian, et al. Hard disk drive failure prediction for mobile edge computing based on an LSTM recurrent neural network[J]. *Mobile information systems*, 2021(2): 8878364.
- [14] YANG Hongzhang, LI Zongzhao, QIANG Huiyuan, et al. ZTE-predictor: disk failure prediction system based on LSTM[C]//2020 50th Annual IEEE-IFIP International Conference on Dependable Systems and Networks-Supplemental Volume. Valencia: IEEE, 2020: 17–20.
- [15] 兰云龙. 基于深度学习的磁盘健康度评估与故障预测技术研究[D]. 杭州: 杭州电子科技大学, 2017.
- LAN Yunlong. Research on disk health evaluation and fault prediction technology based on deep learning[D]. Hangzhou: Hangzhou Dianzi University, 2017.
- [16] 姜少彬, 杜春, 陈浩, 等. 一种硬盘故障预测的非监督对抗学习方法[J]. *西安电子科技大学学报*, 2020, 47(2): 118–125.
- JIANG Shaobin, DU Chun, CHEN Hao, et al. Unsupervised adversarial learning method for hard disk failure prediction[J]. *Journal of Xidian University*, 2020, 47(2): 118–125.
- [17] HU Lihan, HAN Lixin, XU Zhenyuan, et al. A disk fail-

- ure prediction method based on LSTM network due to its individual specificity[J]. *Procedia computer science*, 2020, 176: 791–799.
- [18] WU Jian, YU Haiyang, YANG Zhen, et al. Disk failure prediction with multiple channel convolutional neural network[C]//2021 International Joint Conference on Neural Networks. Shenzhen: IEEE, 2021: 1–8.
- [19] XU Chang, WANG Gang, LIU Xiaoguang, et al. Health status assessment and failure prediction for hard drives with recurrent neural networks[J]. *IEEE transactions on computers*, 2016, 65(11): 3502–3508.
- [20] HAI Qinda, ZHANG Shuangwang, LIU Chang, et al. Hard disk drive failure prediction based on GRU neural network[C]//2022 IEEE/CIC International Conference on Communications in China. Sanshui: IEEE, 2022: 696–701.
- [21] XU Shicheng, XU Xiaolong. ConvTrans-TPS: a convolutional transformer model for disk failure prediction in large-scale network storage systems[C]//2023 26th International Conference on Computer Supported Cooperative Work in Design. Rio de Janeiro: IEEE, 2023: 1318–1323.
- [22] SUN Xiaoyi, CHAKRABARTY K, HUANG Ruirui, et al. System-level hardware failure prediction using deep learning[C]//2019 56th ACM/IEEE Design Automation Conference. Las Vegas: IEEE, 2019: 1–6.
- [23] GUAN Yunchuan, LIU Yu, ZHOU Ke, et al. A disk failure prediction model for multiple issues[J]. *Frontiers of information technology & electronic engineering*, 2023, 24(7): 964–980.
- [24] XIAO Jiang, XIONG Zhuang, WU Song, et al. Disk failure prediction in data centers via online learning[C]//Proceedings of the 47th International Conference on Parallel Processing. Eugene: ACM, 2018: 1–10.
- [25] ZHANG Ji, ZHOU Ke, HUANG Ping, et al. Transfer learning based failure prediction for minority disks in large data centers of heterogeneous disk systems[C]//Proceedings of the 48th International Conference on Parallel Processing. Kyoto: ACM, 2019: 1–10.
- [26] HAN Shujie, LEE P P C, SHEN Zhirong, et al. StreamD-FP: a general stream mining framework for adaptive disk failure prediction[J]. *IEEE transactions on computers*, 2023, 72(2): 520–534.
- [27] 关东海, 赵瑞, 袁伟伟, 等. 一种基于迁移学习和主动学习的硬盘故障预测方法及系统[P]. 中国: CN111949459A, 2020.
- Guan Donghai, ZHAO Rui, YUAN Weiwei, et al. A hard disk failure prediction method and system based on transfer learning and active learning[P]. China: CN111949459A, 2020.
- [28] LU Sidi, LUO Bing, Patel T, et al. Making disk failure predictions SMARTer![C]//Proceedings of the 18th USENIX Conference on File and Storage Technologies. Santa Clara: USENIX Association, 2020: 151–167.
- [29] HAN Shujie, LEE P, XU Fan, et al. An in-depth study of correlated failures in production SSD-based data centers[C]//19th USENIX Conference on File and Storage Technologies. Olivia: USENIX Association, 2021: 417–429.
- [30] WANG Wang, TANG Xuehai, ZHOU Biyu, et al. Improving disk failure detection accuracy via data augmentation[C]//2022 IEEE/ACM 30th International Symposium on Quality of Service. Oslo: IEEE, 2022: 1–10.
- [31] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11531–11539.
- [32] 李倩玉, 王蓓, 金晶, 等. 基于双向 LSTM 卷积网络与注意力机制的自动睡眠分期模型[J]. *智能系统学报*, 2022, 17(3): 523–530.
- LI Qianyu, WANG Bei, JIN Jing, et al. Automatic sleep staging model based on the bi-directional LSTM convolutional network and attention mechanism[J]. *CAAI transactions on intelligent systems*, 2022, 17(3): 523–530.

作者简介:



张铭泉, 副教授, 博士, 主要研究方向为机器学习、计算机体系结构、区块链技术, 发表学术论文 20 余篇。
E-mail: mqzhang@ncepu.edu.cn。



王宝兴, 硕士研究生, 主要研究方向为深度学习和故障检测。E-mail: lifebaoxing@foxmail.com。