

# 一种融合注意力机制的CNN-BiGRU磁盘故障预测方法研究

王艳, 刘亚东, 皮婵娟, 施君豪  
华东交通大学软件学院, 江西 南昌 330013

## 摘要

磁盘作为重要的存储介质,一旦出现故障很可能会导致存储数据丢失,给个人及企业带来难以估量的损失。现有磁盘故障预测模型存在不能很好地平衡磁盘数据样本、未充分利用磁盘数据的时序特性等问题。以Backblaze云存储公司公布的真实磁盘数据为研究对象,提出了一种融合注意力机制的卷积神经网络(CNN)和双向门控循环单元(BiGRU)网络的磁盘故障预测模型。在数据预处理方面,采用负采样与焦点损失函数来平衡正负样本,利用CNN进行特征提取,并结合BiGRU网络来有效地处理时序数据。通过融合注意力机制,能够让模型快速地捕捉更多关键特征信息,将筛选出的特征与数据输入模型进行训练。通过对比其他故障预测模型,本文提出的模型在精确率等4个评价指标上均有1%~7%的性能提升,为提高磁盘存储的可靠性提供了有力的支撑。

## 关键词

注意力机制; 磁盘故障预测; 双向门控循环单元; 卷积神经网络; 焦点损失函数

中图分类号: TP302

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2024060

## *Research on a CNN-BiGRU disk fault prediction method integrating attention mechanism*

WANG Yan, LIU Yadong, PI Chanjuan, SHI Junhao

School of Software, East China Jiaotong University, Nanchang 330013, China

## *Abstract*

Disk, as a crucial storage medium, can result in significant data loss if it malfunctions, causing immeasurable losses for individuals and businesses. Existing models for predicting disk failures have problems such as imbalanced disk data samples and underutilization of the temporal characteristics of the data. In this study, we focused on real disk data provided by the Backblaze cloud storage company and proposed a disk failure prediction model that combines a convolutional neural network (CNN) with a bidirectional gated recurrent unit (BiGRU) network, incorporating an attention mechanism. In terms of data preprocessing, we employed negative sampling and a focal loss function to balance positive and negative samples. Subsequently, we utilized CNN for feature extraction and combined it with BiGRU to effectively handle temporal data. The integration of an attention mechanism enables the model to quickly capture more critical feature

informations. The selected features were then trained with the input data into the model. Compared to other fault prediction models, the proposed model in this paper demonstrates a performance improvement of 1% to 7% on four evaluation indicators, such as precision. This provides a robust support for enhancing disk storage reliability.

### Key words

attention mechanism, disk failure prediction, convolutional neural network, bidirectional gated recurrent unit, focal loss function

## 0 引言

随着物联网、区块链、云计算等技术的兴起,人类在各种互联网活动中产生的数据量激增,磁盘作为计算机系统中重要的存储介质,在大数据时代起着重要的作用。然而,由于磁盘长时间高强度运行、频繁进行读写操作,磁盘故障成为不容忽视的问题。磁盘故障可能导致数据丢失、系统崩溃以及出现信用危机,对企业及个人都会带来难以估量的损失,因此对磁盘故障进行预测和及时处理显得尤为重要。本文提出了一种新的磁盘故障预测模型,将注意力机制融入卷积神经网络(convolution neural network, CNN)和双向门控循环单元(bidirectional gated recurrent unit, BiGRU)网络中,提高模型对关键特征的捕捉能力,更准确地预测磁盘的健康状态。同时,本文采用负采样与焦点损失函数相结合的方式平衡正负样本,解决了数据不平衡问题,提高了模型的预测性能。本文主要贡献如下。

- 构建了磁盘故障预测模型。本文提出了融合注意力机制的CNN-BiGRU磁盘故障预测模型,弥补了其他深度学习故障预测模型在预测性能上的不足。

- 解决了数据样本不平衡问题。本文引入焦点损失函数来处理磁盘正负样本数据极度不平衡的问题,提升了模型的预测性能。

- 进行了实验结果分析。与其他主流预测模型进行对比,本文模型在准确率(accuracy)、精确率(precision)、召回率(recall)及F1-score 4个评价指标上均有1%~7%的性能提升。

## 1 相关工作

为了保障磁盘数据存储的可靠性,国内外许多学者与研究团队致力于提出有效的磁盘故障预测方法,目前主流的磁盘故障预测方法主要分为两类<sup>[1]</sup>。

(1) 基于统计学的磁盘故障预测方法

主要有阈值法和距离异常检测法,这些方法通常采用磁盘自身SMART属性阈值及历史故障数据来确定磁盘是否发生故障。Hughes等<sup>[2]</sup>运用磁盘历史运行数据进行分析建模,提出两种统计学方法,并采用秩和检验(rank sum test)对磁盘故障做出预测,并在实验数据集上实现了60%的故障检测率(false discovery rate, FDR)和0.2%的误报率(false alarm rate, FAR)。Wang等<sup>[3]</sup>基于异常检测法提出了一种马氏距离两步参数化方法来建立磁盘故障预测模型,采用基于滑动窗口的广义似然比来预测磁盘故障,在实验数据集上实现了67%的FDR。Huang等<sup>[4]</sup>提出了一种基于统计学的欧氏距离预测方法,旨在通过对比磁盘的磁头故障和扇区故障与故障磁盘的欧氏距离来判别新磁盘样本的

健康情况。由于存在磁盘厂商在阈值设置方面较保守等原因,基于统计学的故障预测方法在当下大规模异构磁盘环境下的预测效果并不理想。

(2) 基于机器学习和深度学习的磁盘故障预测方法

国防科技大学董勇等<sup>[6]</sup>对多种机器学习算法进行了比较,其中反向传播神经网络(back propagation neural network, BPNN)的FDR稳定在93%左右,但其FAR的稳定性相对较差,在0.1%和1%之间浮动。对于决策树(decision tree, DT)、贝叶斯(Bayes)及支持向量机(support vector machine, SVM)等预测模型,其FDR分别为98%、96%及97%左右,由于SVM的FDR对时间窗口的变化较敏感,所以实际的FDR保持在67%至97%,但当时间窗口大于24 h时,该模型的FDR也会呈现稳定趋势。对于以上模型的误报率,SVM的表现相对较好,该模型的FAR仅有0.1%,决策树和贝叶斯方法的FAR分别为0.3和0.8%,性能略差于SVM。在误报率方面,BPNN的稳定性相对较差,其FAR在0.1%至1%浮动。Yang等<sup>[6]</sup>采用逻辑回归(logistic regression, LR)方法构建了一种磁盘故障预测模型,经实验论证分析,该模型获得了97.8%的FDR和0.3%的FAR。Shen等<sup>[7]</sup>提出了一种随机森林(random forest, RF)故障预测模型,并且经实验论证,该模型的FDR和FAR分别为97%和0.02%。在深度学习方面,Xu等<sup>[8]</sup>提出了一种基于循环神经网络(recurrent neural network, RNN)的磁盘故障预测模型,利用RNN来更好地捕捉磁盘数据集的时序特性,从而获得良好的预测效果。Basak等<sup>[9]</sup>开发了一个基于长短期记忆(long short-term memory, LSTM)网络的磁盘故障预测框架,经过论证,该模型的预测效果良好且可以提前

7天预测出即将出现故障的磁盘。Sun等<sup>[10]</sup>融合SMART属性的时序特性,建立了一种基于卷积神经网络(convolutional neural network, CNN)的故障预测模型。Lu等<sup>[11]</sup>充分考虑了磁盘的服务器属性与位置属性,搭建了CNN-LSTM故障预测模型,并对比其他传统的预测模型,取得了较理想的预测效果。Jiang等<sup>[12]</sup>采用生成式对抗网络(generative adversarial network, GAN),将SMART属性数据转化为二维数据,磁盘故障预测问题也就被转化为异常检测问题,最后将模型微调至理想效果。

## 2 数据处理与特征选择

### 2.1 实验数据集来源

Backblaze是一家专注于为个人用户和企业用户提供可靠、安全、高效的数据备份和存储解决方案的公司,每年会发布其数据中心中大量硬盘的使用情况和故障信息的数据集,包括硬盘的型号、序列号、制造商、容量、运行时间、温度、读写错误次数、故障日期等,为学术界、研究人员以及硬盘制造商提供了可持续更新的数据来源。本文采用Backblaze云存储商公布的2023年第一季度磁盘数据作为实验数据集,其中部分故障率较高的磁盘信息见数据表1。

本文将故障磁盘标记为正样本,将健康磁盘标记为负样本。由表1可知,每家磁盘制造商的故障磁盘数量都很少且故障率较低,因此在实验数据集中正样本较少,正负样本极度不平衡。为了减小样本不平衡问题对模型预测性能的影响,本文选择故障比例较高的ST4000DM000型号和ST12000NM0008磁盘型号作为实验数据集,分别记为ST1和ST2。

表1 部分故障较高的磁盘信息数据

磁盘制造商/型号	磁盘总数/个	故障数量/个	磁盘故障比例
Seagate (希捷)/ST16000NM001G	21 910	31	0.14%
Toshiba (东芝)/MG07ACA14TA	38 156	109	0.29%
Seagate (希捷)/ST12000NM0008	19 749	120	0.61%
Seagate (希捷)/ST8000NM0055	14 331	132	0.92%
Seagate (希捷)/ST4000DM000	18 070	170	0.94%

## 2.2 数据预处理

在数据预处理方面, Backblaze公司公布的SMART属性数据集包含原始值和标准化值两类。本文采用时间序列差分<sup>[13]</sup>来对数据进行预处理。时间序列差分是常见的数据预处理技术, 用于将时间序列数据转换成平稳的序列, 以便更好地应用于时间序列的分析和建模。时间序列差分通过计算当前时刻的观测值与前一个时刻的观测值之间的差异, 得到新的序列值。一阶差分可以用下式表示:

$$\Delta X_t = X_t - X_{t-1} \quad (1)$$

其中,  $\Delta X_t$  表示差分值,  $X_t$  表示当前数据点,  $X_{t-1}$  表示前一数据点。

一阶差分操作可以消除时间序列中的线性趋势, 使序列更平稳, 如果时间序列仍不平稳, 可以继续进行多阶差分操作, 直到得到平稳的序列。

## 2.3 平衡样本数据

在平衡正负样本方面, 本文采用了负采样与焦点损失函数 (focal loss)<sup>[14]</sup>。焦点损失函数是一种用于解决类别不平衡问题的损失函数, 特别适用于二分类问题, 旨在增强模型对少数类样本的学习能力, 从而改善在类别不平衡情况下的分类性能, 其计算式如下:

$$\text{focalloss}(p_t) = -(1 - p_t)^\gamma \cdot \log(p_t) \quad (2)$$

其中, 对于二分类问题,  $p_t$  是模型输出的概率值,  $\gamma$  是焦点参数, 同时也是放缩因子, 控制焦点损失函数的损失程度。

由于健康磁盘与故障磁盘的数量极度不平衡, 首先需要进行负采样来平衡正负样本的数量。对于每个正样本, 从负样本中随机采样一定数量的负样本作为训练样本, 然后构造新样本集。在完成负采样后, 将正样本和负样本组合成一个新的样本集, 该新样本集中, 正负样本的数量应该接近平衡, 本文将正负样本比例设置为1:5, 接着将焦点损失函数作为优化目标函数。焦点损失函数会增大少数类别样本的权重, 通过焦点因子使模型更加关注少数类样本, 从而改善模型在数据类别不平衡问题上的性能。两种方法结合使用可以更好地平衡正负样本的数量, 最终提升模型的预测性能。本文选取1 000个磁盘样本数据, 将训练数据集和测试数据集以8:2的比例进行划分。

## 2.4 特征提取与特征选择

在特征提取方面, 本文采用CNN进行特征提取。CNN在模型的前半部分用于从原始输入数据中提取关键特征。通过多层卷积和激活函数, CNN能够捕捉数据的局部模式和空间结构, 从而将输入数据转换为更具代表性和更高层次的特征来表示。

Attention机制<sup>[15]</sup>用于在模型的后半部分对CNN提取的特征进行加权处理,通过对特征进行加权,能够让模型更加关注重要的特征信息,忽略不重要的部分,从而提高模型的预测性能和泛化能力。在本文提出的CNN-BiGRU-Attention模型中,Attention机制能够快速捕捉更多关键的特征信息,尤其是针对磁盘故障预测这类复杂的时序数据,Attention机制能够帮助模型更有效地识别关键特征,提高预测性能。

本文的实验数据集有174个SMART属性特征值,由于数据集中存在大量空值,首先,在将空值进行剔除并将缺失值进行填充后,还剩48个属性特征值,本文筛选出20个特征作为模型的输入特征,见表2。

### 3 模型与方法

目前,磁盘故障预测领域采用的主流深度学习方法有RNN(recurrent neural network)、LSTM、GRU(gated recurrent unit)<sup>[16]</sup>及CNN-LSTM等。RNN是专门用于处理序列数据的神经网络,它通过在每个时间步使用相同的权重来处理

序列数据,并且通过时间反向传播来捕捉数据的时序依赖关系。但在传统的RNN中,由于时间步的反向传播,很容易出现梯度消失或梯度爆炸问题,导致训练效果不理想。LSTM是RNN的一个变体,它通过添加输入门、遗忘门和输出门来解决梯度消失和梯度爆炸问题,更好地捕捉长期依赖关系。GRU是RNN的另一个变体,它和LSTM很相似但参数更少,计算效率相对更高。虽然GRU的结构相较LSTM更加简单、收敛速度更快,但LSTM和GRU在处理长序列数据时仍可能出现梯度消失问题。CNN-LSTM能更好地捕捉数据的空间特征与时序依赖关系,但其参数量仍较大,且在处理小规模数据或噪声较多的数据时容易发生过拟合现象。

本文的实验数据集以天为单位对磁盘SMART属性数据进行采集,因此可以将磁盘数据集视为一组时间序列数据。注意力机制更加聚焦于重要的输入数据,从而提高深度学习网络模型的性能。例如在序列数据中,自然语言文本、音频信号、时间序列等,注意力机制根据上下文和语境来决定哪些部分对于当前任务更重要。在磁盘故障预测的任务中,注意力机制的主要

表2 最终筛选出的20个特征

属性ID	变量解释	属性ID	变量解释
smart_1_raw	原始读取错误率	smart_188_raw	命令超时
smart_3_raw	启动时间	smart_189_raw	磁头写入高度
smart_4_raw	电机启停次数	smart_191_raw	加速度错误率
smart_5_raw	重映射扇区数	smart_192_raw	电源关闭磁头收回计数
smart_7_raw	寻道错误率	smart_193_raw	磁头升降次数
smart_9_raw	通电时间	smart_194_raw	温度计数
smart_10_raw	旋转重试次数	smart_197_raw	当前待定扇区数
smart_12_raw	设备周期计数	smart_198_raw	不可纠正扇区数
smart_184_raw	端对端出错	smart_240_raw	传输错误率
smart_187_raw	报告不可纠正错	smart_241_raw	LBA写入总数

作用如下。

- 时间序列数据分析：磁盘运行时生成的SMART数据是时间序列数据，包括磁盘温度、读写速度、错误计数等，在预测磁盘故障时，注意力机制可以帮助模型找到最重要的时间点或特征。

- 特征关联学习：注意力机制可以用于学习不同特征之间的关联性，比如磁盘温度和读写速度之间的关系，帮助模型更好地捕捉潜在的关系模式。

相较于GRU，双向门控循环单元（bidirectional gated recurrent unit, BiGRU）具有双向结构，可以更全面地捕捉时间序列数据中的依赖关系，而GRU只考虑了过去的信息，对于一些长期依赖关系的捕捉可能表现效果不如BiGRU。其次，BiGRU和GRU都使用了门控单元来控制信息的流动，以解决梯度消失和梯度爆炸问题。在结构方面，BiGRU中具有更新门和重置门两个单元，而GRU中只有一个更新门。另外，

BiGRU在处理序列数据时，相较于传统的LSTM网络有更好的梯度传播和信息捕捉能力，且参数更少，计算效率更高。

本文建立了一种融合注意力机制的CNN-BiGRU磁盘故障预测模型，故障预测流程如图1所示。

本文使用的磁盘故障预测模型主要由3个模块构成，分别是卷积神经网络（CNN）模块、双向门控循环单元（BiGRU）模块和注意力（Attention）模块，CNN-BiGRU-Attention模型结构如图2所示。其中，CNN模块由5层一维卷积层、5层Mish激活函数、5层批归一化层和1层下采样层组成，通过使用Mish激活函数和归一化操作来进行特征提取。BiGRU是一种适用于序列数据处理的循环神经网络结构，能够同时考虑历史和未来的信息。BiGRU模型相较于传统GRU模型，结合了前向GRU和后向GRU，能够同时从过去和未来的信息中提取特征，有助于更好地捕捉序列数据中的模式。BiGRU模块主要由双向门控循环单元和融合层组成，双向门控循环单元主要用于处理序列数据中的时序信息并进行特征提取，融合层用于将多个特征进行融合，以增强模型的表达能力，从而有助于模型更好地理解输入数据并达到更好的预测性能。

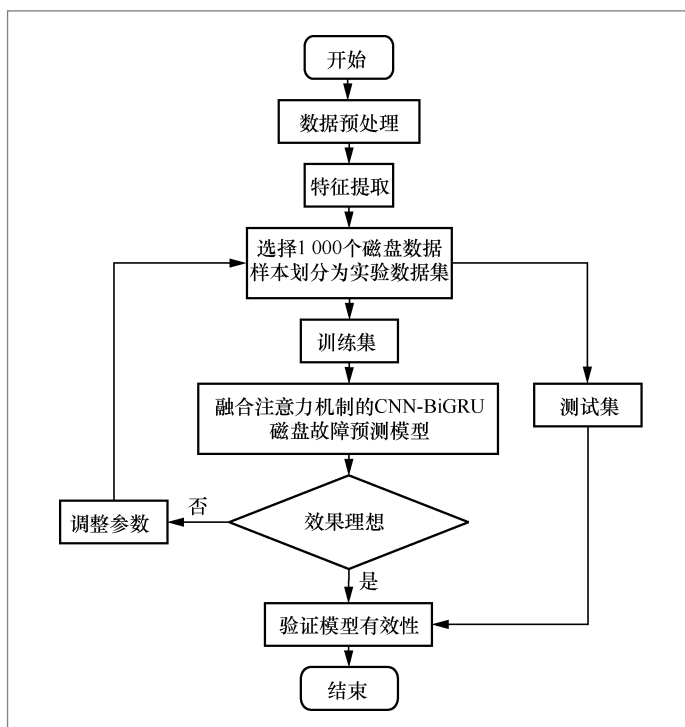


图1 故障预测流程

## 4 实验结果与分析

### 4.1 实验平台

本文的实验平台采用Windows11操作系统，处理器型号为AMD Ryzen9 7945HX，显卡型号为NVIDIA GeForce RTX4060，编程语言采用Python 3.9.13，深度学习框架为Pytorch 2.0.0+cu118，机器学习框架为Scikit-learn 1.3.0。

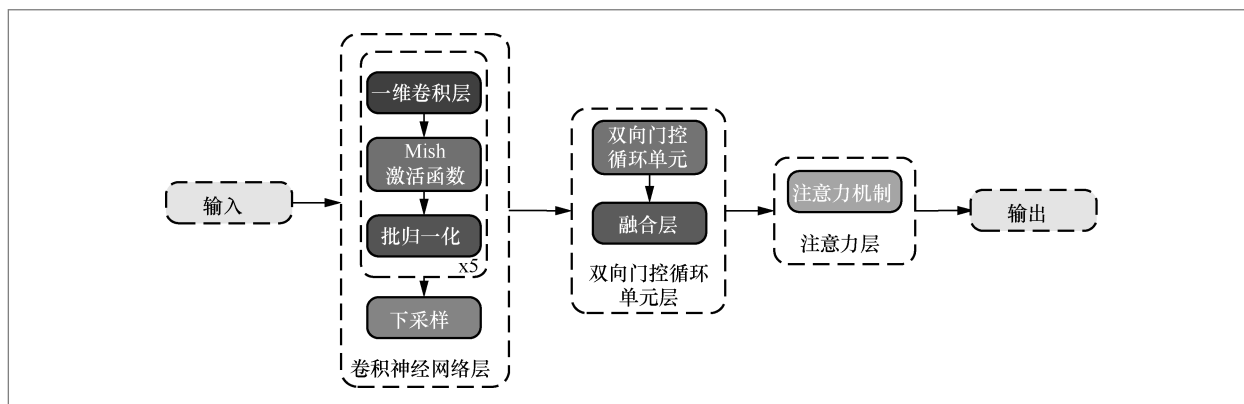


图2 CNN-BiGRU-Attention 模型结构

## 4.2 评价指标

对于典型的二分类问题通常有以下4种结果：TP (true positive)、TN (true negative)、FP (false positive) 和FN (false negative)，用于评估二分类模型在预测中的表现。其中，TP表示模型将正类样本正确地预测为正类的数量，即模型真正地识别了正例；TN表示模型将负类样本正确地预测为负类的数量，即模型真正地识别了负例；FP表示模型将负类样本错误地预测为正类的数量，即模型将负例误判为正例；FN表示模型将正类样本错误地预测为负类的数量，即模型将正例误判为负例。

准确率 (accuracy)、精确率 (precision)、召回率 (recall) 及F1-score常被用作评价分类模型性能的指标，它们用于衡量模型在不同方面的表现。为了评估本文模型预测的性能，统一采用上述指标作为本文模型性能的评价指标，以下是各项指标的相关介绍及计算式。

### (1) 准确率

表示模型正确预测的样本数量占总样本数量的比例，准确率衡量了整体预测的正确性，计算式如下：

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

### (2) 精确率

又称查准率，表示在模型预测为正例的样本中实际为正例的比例，精确率描述模型对于正例预测的准确性，计算式如下：

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

### (3) 召回率 (recall)

又称查全率，表示在实际为正例的样本中模型预测为正例的比例，召回率关注于模型对于真正正例的识别能力，计算式如下：

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

### (4) F1-score

F1-score是一个结合精确率和召回率的综合评价指标，综合考虑了精确率和召回率的平衡，计算式如下：

$$\text{F1-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

## 4.3 实验结果分析

### (1) 模型对比实验分析

本文将训练数据集和测试数据集以

8:2的比例切分,模型超参数Epochs设置为100,损失函数权重参数Gamma设置为5,学习率lr设置为 $1 \times 10^{-4}$ ,5层一维卷积的输入通道数为1,卷积核数分别为8、16、32、64、128,卷积核大小分别为3、3、3、7、9,步长为1。本文共选取8种其他磁盘预测模型与本文提出的模型进行了对比,各模型参数见表3。在ST1数据集下每种模型的预测效果见表4,将结果可视化如图3所示。

由图3可知, Bayes预测方法的性能相

表3 模型主要参数配置

模型	模型参数设置
Bayes	priors=None, var_smoothing=1e-9
MLP	hidden_layer_sizes=(50, ), activation=relu, solver=adam, alpha=0.0001, batch_size=auto
GBoost	learning_rate=0.08, n_estimators=200, subsample=0.5, random_state=None, verbose=0
SVM	kernel=linear, degree=1, gamma=scale, coef0=0.0, tol=0.001, catch_size=50, max_iter=-1
CNN-LSTM	input_size=20, hidden_size=2, num_layers=2, dropout=0.2
CNN-BiLSTM	input_size=20, hidden_size=2, num_layers=2, bidirectional=True, dropout=0.2
CNN-BiGRU	input_size=20, hidden_size=2, num_layers=2, bidirectional=True, dropout=0.2

表4 ST1数据集下对比实验结果

模型	accuracy	precision	recall	F1-score
Bayes	0.87	0.90	0.87	0.88
MLP	0.91	0.90	0.90	0.86
GBoost	0.91	0.90	0.90	0.86
SVM	0.95	0.94	0.94	0.94
CNN	0.87	0.92	0.86	0.87
CNN-LSTM	0.88	0.93	0.88	0.89
CNN-BiLSTM	0.93	0.94	0.94	0.93
CNN-BiGRU	<b>0.95</b>	0.94	0.93	0.94
本文模型	0.94	<b>0.96</b>	<b>0.94</b>	<b>0.95</b>

对较差,传统CNN预测方法的性能也不太理想,而采用能充分捕捉磁盘时序性的模型性能普遍较高。进一步分析也可以发现,具有双向结构的模型(BiLSTM、BiGRU)性能会比传统模型(LSTM、GRU)性能提升更加明显,究其原因总结如下。首先,双向模型分别从前向和后向处理输入数据,可以同时获取上下文信息,更好地捕捉序列的长期依赖关系。其次,在传统LSTM和GRU中,很容易出现梯度消失或爆炸问题,而双向结构网络可以有效解决这一问题。在准确率方面,SVM和CNN-BiGRU可以达到95%,SVM作为一种经典的分类模型,可以将低维的输入数据映射到高维特征空间中,从而在高维空间中寻找最优的超平面来完成分类任务。

在其他3个评价指标中,本文提出的模型精确率为0.96、召回率为0.94、F1-score为0.95,对比其他模型性能提升了1%至7%。这是因为双向结构的GRU可以较好地处理长时间序列问题。GRU网络结构相较于LSTM网络,GRU只有两个门(更新门和重置门),而LSTM具有3个门(输入门、遗忘门和输出门),因此在计算效率上GRU会更高。

GRU有较少的门数且有助于解决梯度消失的问题,这也就使BiGRU在处理某些特定时间序列任务时更容易训练和优化,更加容易捕捉序列的长期依赖关系。对比发现加入注意力机制的BiGRU在4个评价指标上略高于传统BiGRU。经过上述分析及实验结果,可以证实本文提出的模型优越性及合理性。

在ST2数据集下的每种模型预测效果如表5和图4所示。从表5、图4中可以发现,在此数据集下,传统CNN预测模型预测的性能较差,准确率和召回率仅为0.56。还可以发现在此数据集下机器学习分类模型的性能普遍略好于深度学习模型,究其

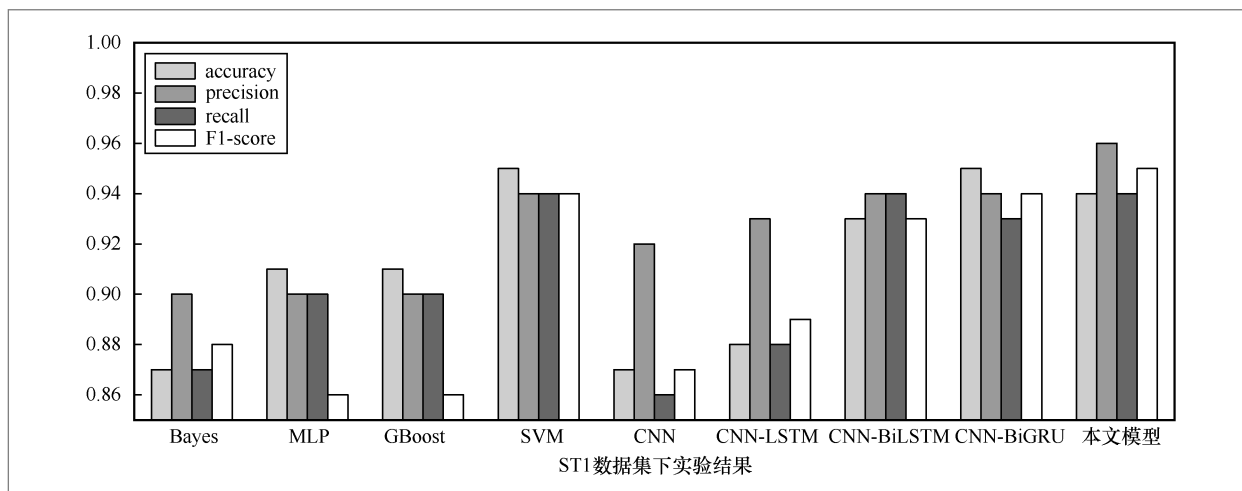


图3 ST1数据集下预测模型对比

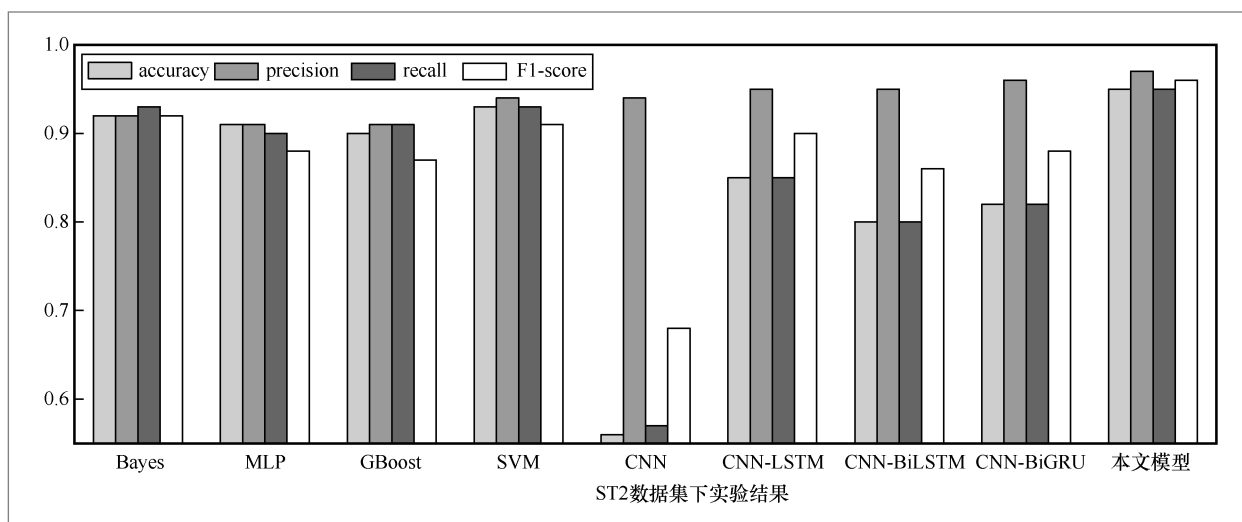


图4 ST2数据集下预测模型对比

原因是不同数据集可能存在不同的特征分布和噪声等。深度学习模型会更容易受到噪声的影响,且ST2数据集相较于ST1数据集样本规模更小,在较小数据集上机器学习模型更具优势。在精确率和F1-score上,本文提出的模型明显优于其他模型,分别为0.97和0.96,进一步可以证明在引入注意力机制后,会进一步增强模型捕捉特征的能力,从而使模型性能更优。在准确率方面,本文提出的模型与SVM表现更优,分别达到了0.95和0.93,远高于传统CNN模型。在召回率方面,本文提出的模

表5 ST2数据集下对比实验结果

模型	accuracy	precision	recall	F1-score
Bayes	0.92	0.92	0.93	0.92
MLP	0.91	0.91	0.90	0.88
GBoost	0.90	0.91	0.91	0.87
SVM	0.93	0.94	0.93	0.91
CNN	0.56	0.94	0.57	0.68
CNN-LSTM	0.85	0.95	0.85	0.90
CNN-BiLSTM	0.80	0.95	0.80	0.86
CNN-BiGRU	0.82	0.96	0.82	0.88
本文模型	0.95	0.97	0.95	0.96

型达到了0.95,主流深度学习故障预测模型CNN、CNN-LSTM仅有0.57和0.85,因此在召回率方面本文提出的模型性能更优。经实验结果分析,验证了本文提出的故障预测模型的有效性及其合理性。

#### (2) Loss值变化曲线分析

Loss值是衡量模型预测性能的重要指标,其变化曲线反映了模型在训练过程中的优化效果。本文提出的模型的ST1数据集训练集Loss曲线如图5所示,横轴表示训练轮次(Epoch),纵轴表示Loss值。

随着Epoch的增加, Loss值逐渐减小,表明模型的预测性能在优化过程中得到了不断改善。从Loss值变化曲线可以看出,在训练起始阶段Loss值较高,

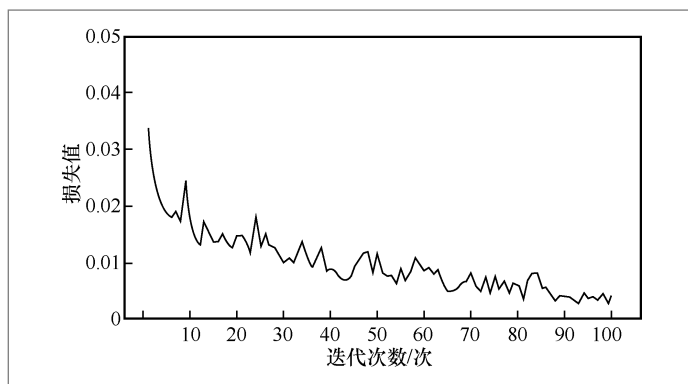


图5 CNN-BiGRU-Attention 模型 Loss 值变化曲线

表6 Attention 组件对模型效果的影响

模型	accuracy	precision	recall	F1-score
CNN	0.87	0.92	0.88	0.90
CNN-Attention	0.92	0.94	0.92	0.93
CNN-LSTM	0.88	0.93	0.88	0.90
CNN-LSTM-Attention	0.90	0.95	0.90	0.91
CNN-GRU	0.86	0.93	0.86	0.89
CNN-GRU-Attention	0.89	0.94	0.89	0.90
CNN-BiGRU	0.95	0.94	0.93	0.94
本文模型	0.94	0.96	0.94	0.95

在Epoch从8到9时震荡幅度最大,约为0.02,这表明模型的预测性能尚未得到充分发挥,但随着训练的进行,模型逐渐学习到数据的特征和规律, Loss值开始下降。在这个阶段,模型正在逐步拟合训练数据,提高了对训练样本的拟合能力。

随着训练的进行,在Epoch为11至88时, Loss值逐渐趋于稳定,并在一个较小的范围内波动。这表明模型已经较好地学习了数据的特征,并且对训练样本和验证样本都表现出较好的预测能力。在这个阶段,模型已经取得了较好的收敛效果。

当Epoch从88开始时,训练集Loss值变化逐渐趋于平稳且波动幅度小,可能是训练数据的噪声或模型的复杂性导致的,也可能是模型在局部最优点附近进行微调导致的。在实验中,通过合理的超参数设置和模型结构调优,取得了较好的Loss值变化曲线,证明了所提出的CNN-BiGRU-Attention模型在磁盘故障预测任务中的有效性和性能优势。

## 4.4 实验结果分析

### (1) 组件消融实验

为了进一步评估Attention组件对模型性能的影响,本文采用4组组件消融对比实验来验证模型的优越性及合理性,使用ST1数据集,分别采用CNN、CNN-LSTM、CNN-GRU、CNN-BiGRU模型,在各个模型中通过添加和去除Attention组件来验证有效性,结果见表6。

将表6结果可视化,如图6至图9所示。以4个评价指标及数值分别为横纵坐标绘制曲线,通过分析图表可以得出在基础模型上添加了Attention模块后可以有效提升模型的预测性能,性能增幅约为0.01至0.07。对此总结原因如下。

首先,注意力机制让模型更加关注重

要的特征,而忽略不重要的特征,通过学习特征的权重,注意力机制能够有效地选择那些对于当前任务最有用的特征,从而提升模型的性能;其次,磁盘数据集属于时间序列数据,对于传统的处理时序问题的模型,例如RNN、LSTM、GRU,可能会出现梯度消失或梯度爆炸的情况,导致难以捕捉长期序列的依赖关系。而注意力机制能够在序列中自适应地加权考虑每个时间步的信息,更好地捕捉序列的长期依赖关系;最后,在模型训练过程中注意力机制可以减少模型对噪声和冗余信息的敏感性,更好地提升模型的预测性能。综上,注意力机制能够使模型更加灵活和准确地处理复杂的数据关系,提升模型的预测性能。

## (2) 模型结构消融实验

为了进一步分析模型结构对于模型预测性能的影响,本文采用CNN-LSTM和CNN-GRU两组模型结构消融实验来验证双向结构的有效性及其合理性,使用ST1数据集,在各个模型中通过更改模型结构双向性来验证有效性,结果见表7。

模型结构效果对比如图10、图11所示,在实验中,本文分别设计了单向结构和双向结构的模型进行对比,以研究双向结构在磁盘故障预测中的影响。首先,保持模型的卷积神经网络部分不变,但将双向门控循环单元(BiGRU)替换为普通的单向门控循环单元(GRU);其次,保持模型的卷积神经网络部分不变,并继续采用BiGRU作为模型的序列建模部分。实验结果分析如下。在单向结构实验组中,模型在准确率、精确率、召回率和F1-score等指标上表现良好,而在双向结构实验组结果显示,模型在准确率、精确率、召回率和F1-score等指标上相较于单向结构实验组有明显的提升。这是因为相比于单向结构,双向结构的优势在于能够更好地处理序列数据中的长期依赖关系,避免了梯度消失,

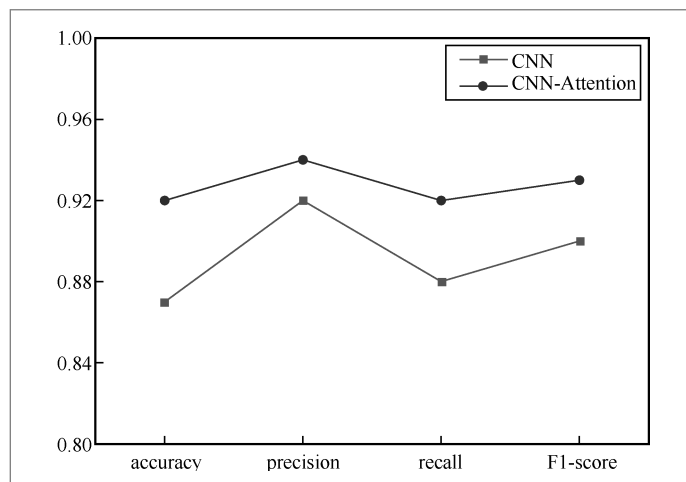


图6 Attention 组件对 CNN 模型影响

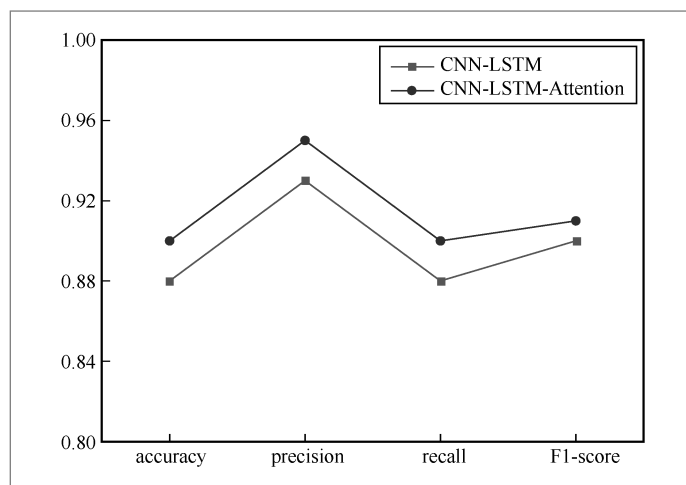


图7 Attention 组件对 CNN-LSTM 模型影响

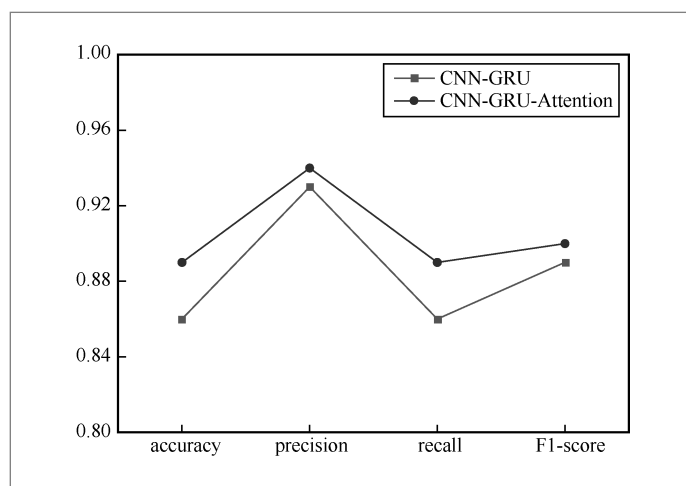


图8 Attention 组件对 CNN-GRU 模型影响

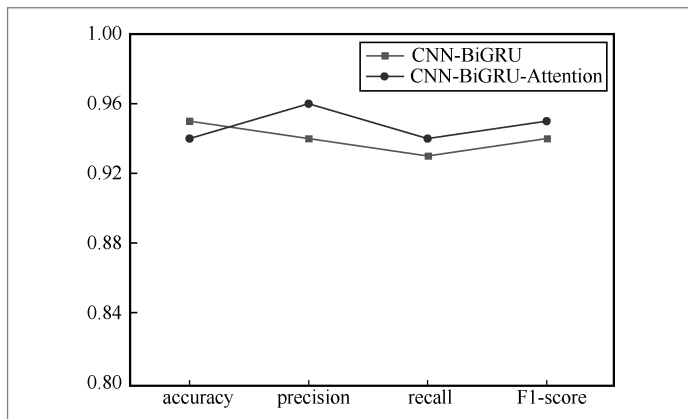


图9 Attention 组件对 CNN-BiGRU 模型的影响

表7 双向结构对模型性能的影响

模型	accuracy	precision	recall	F1-score
CNN-LSTM	0.88	0.93	0.88	0.90
CNN-BiLSTM	0.94	0.94	0.94	0.93
CNN-GRU	0.86	0.93	0.86	0.89
CNN-BiGRU	0.95	0.94	0.93	0.94

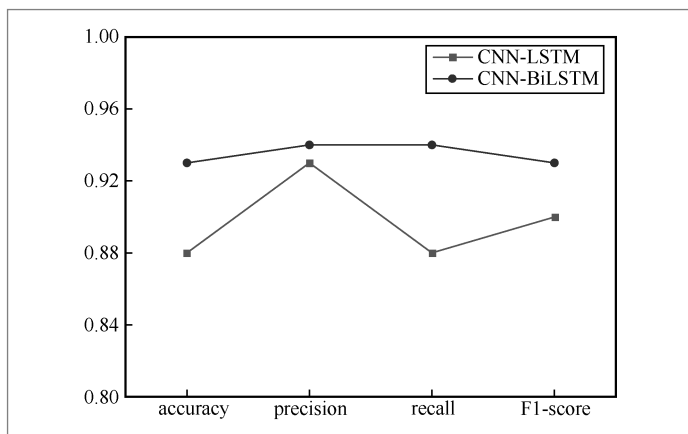


图10 单双向 LSTM 结构对比

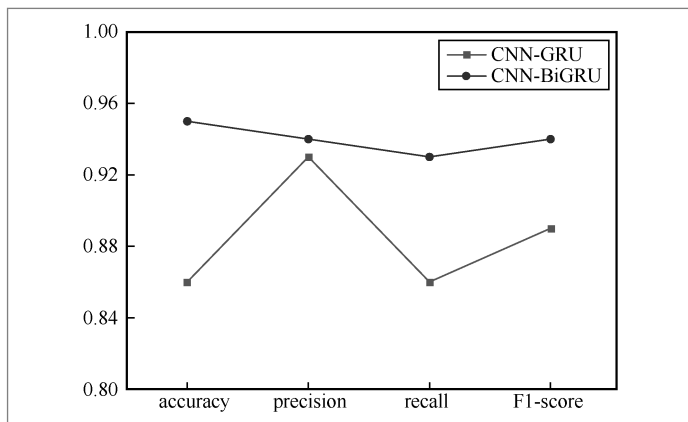


图11 单双向 GRU 结构对比

并且能够更好地捕捉序列数据的上下文信息,从而提升了磁盘故障预测模型的性能。综上,双向结构网络在磁盘故障预测模型中表现出优势,为进一步优化磁盘故障预测模型提供了相关参考。

## 5 总结与展望

本文提出了一种基于注意力机制的 CNN-BiGRU 模型用于磁盘故障预测,并对其进行了详细的实验评估。通过在 Backblaze 公司公布的真实磁盘数据集上进行实验,本文验证了该模型的有效性和优越性。首先,采用负采样及焦点损失函数来平衡正负样本,解决了数据集中样本不平衡问题,提高了模型的性能;其次在模型结构方面,CNN 用于提取特征,并加入 BiGRU 来建模时序数据,充分考虑了序列数据中的长期依赖关系;同时,本文引入了注意力机制,使模型能够自适应地选择关键特征,进一步提高了预测性能;最后实验结果表明,本文提出的 CNN-BiGRU-Attention 模型在磁盘故障预测任务中取得了优秀的性能。与传统的深度学习模型相比,本文提出的模型在准确率、精确率、召回率和 F1-score 等指标上都有所提升,证明了该模型的有效性。

然而,本研究还存在一定的局限性。例如,本文使用了特定的数据集进行实验,可能存在数据分布的偏差问题。此外,模型的参数设置和网络结构仍有进一步优化的空间,可以通过更多的实验来进一步提升模型的性能。在未来的研究中,需要进一步考虑引入导致磁盘故障的其他因素,例如磁盘机架属性,数据中心地理位置温度、湿度等,进一步提高磁盘故障模型的预测性能。

## 参考文献:

- [1] 彭福康. 面向云数据中心磁盘时序数据的故障预测方法研究[D]. 郑州: 郑州大学, 2021.  
PENG F K. Research on failure prediction method for disk time series data in cloud data center[D]. Zhengzhou: Zhengzhou University, 2021.
- [2] HUGHES G F, MURRAY J F, KREUTZ-DELGADO K, et al. Improved disk-drive failure warnings[J]. IEEE Transactions on Reliability, 2002, 51(3): 350-357.
- [3] WANG Y, MA E W M, CHOW T W S, et al. A two-step parametric method for failure prediction in hard disk drives[J]. IEEE Transactions on Industrial Informatics, 2014, 10(1): 419-430.
- [4] HUANG S, FU S, ZHANG Q, et al. Characterizing disk failures with quantified disk degradation signatures: an early experience[C]//Proceedings of the 2015 IEEE International Symposium on Workload Characterization. Piscataway: IEEE Press, 2015: 150-159.
- [5] 董勇, 蒋艳凤, 卢宇彤, 等. 面向磁盘故障预测的机器学习方法比较[J]. 计算机工程与科学, 2015, 37(12): 2200-2207.  
DONG Y, JIANG Y H, LU Y T, et al. Comparison of machine learning methods for disk failure prediction[J]. Computer Engineering & Science, 2015, 37(12): 2200-2207.
- [6] YANG W J, HU D M, LIU Y L, et al. Hard drive failure prediction using big data[C]//Proceedings of the 2015 IEEE 34th Symposium on Reliable Distributed Systems Workshop (SRDSW). Piscataway: IEEE Press, 2015: 13-18.
- [7] SHEN J, WAN J, LIM S J, et al. Random-forest-based failure prediction for hard disk drives[J]. International Journal of Distributed Sensor Networks, 2018, 14(11).
- [8] XU C, WANG G, LIU X G, 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.
- [9] BASAK S, SENGUPTA S, DUBEY A. Mechanisms for integrated feature normalization and remaining useful life estimation using LSTMs applied to hard-disks[C]//Proceedings of the 2019 IEEE International Conference on Smart Computing (SMARTCOMP). Piscataway: IEEE Press, 2019: 208-216.
- [10] SUN X Y, CHAKRABARTY K, HUANG R R, et al. System-level hardware failure prediction using deep learning[C]//Proceedings of the 2019 56th ACM/IEEE Design Automation Conference (DAC). Piscataway: IEEE Press, 2019: 1-6.
- [11] LU S, LUO B, PATEL T, et al. Making disk failure predictions SMARTer![C]//Proceedings of 18th USENIX Conference on File and Storage Technologies (FAST 20). Hyatt Regency Santa Clara: USENIX, 2020: 151-167.
- [12] JIANG T M, ZENG J F, ZHOU K, et al. Lifelong disk failure prediction via GAN-based anomaly detection[C]//Proceedings of the 2019 IEEE 37th International Conference on Computer Design (ICCD). Piscataway: IEEE Press, 2019: 199-207.
- [13] 贾彦锋, 朱新慧, 叶家彬, 等. GNSS坐标非线性变化的差分长短时记忆网络预测[J]. 测绘科学, 2022, 47(10): 89-95.  
JIA Y F, ZHU X H, YE J B, et al. Prediction of GNSS coordinate nonlinear variations using difference method and long short term memory[J]. Science of Surveying and Mapping, 2022, 47(10): 89-95.
- [14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2980-2988.

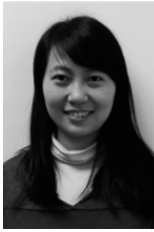
[15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB]. arXiv preprint, 2017, arXiv: 1706.03762.

[16] 李港, 李莉, 林国义, 等. 硬盘故障预测模型的建立与实现[J]. 控制工程, 2022, 29(10):

1788-1792.

LI G, LI L, LIN G Y, et al. Establishment and realization of hard disk failure prediction model[J]. Control Engineering of China, 2022, 29(10): 1788-1792.

### 作者简介



王艳 (1982- ), 女, 博士, 华东交通大学软件学院副教授, 主要研究方向为纠删码、分布式存储系统、数据容错技术。



刘亚东 (1997- ), 男, 华东交通大学软件学院硕士生, 主要研究方向为分布式存储系统、数据容错技术。



皮婵娟 (1999- ), 女, 华东交通大学软件学院硕士生, 主要研究方向为纠删码、分布式存储、分布式计算。



施君豪 (1999- ), 华东交通大学软件学院硕士生, 主要研究方向为分布式存储与编码。

收稿日期: 2024-02-22

通讯作者: 王艳, wangyann@189.cn

基金项目: 国家自然科学基金项目 (No.61962020); 上海市智能信息处理重点实验室开放基金项目 (No.IIPL201910); 江西省教育厅项目 (No.GJJ2200640)

Foundation Items: The National Natural Science Foundation of China (No.61962020), Open Research Program of Shanghai Key Lab of Intelligent Information Processing (No.IIPL201910), Jiangxi Education Department Program (No.GJJ2200640)