

# 一个针对多种问题的磁盘故障预测模型

关云川<sup>1</sup>, 刘渝<sup>2</sup>, 周可<sup>1</sup>, 李强<sup>3</sup>, 王团结<sup>3</sup>, 李辉<sup>3</sup>

<sup>1</sup>华中科技大学武汉光电国家研究中心, 中国武汉市, 430074

<sup>2</sup>华中科技大学计算机科学与技术学院, 中国武汉市, 430074

<sup>3</sup>浪潮电子信息产业股份有限公司, 中国北京市, 250000

**摘要:** 磁盘故障预测方法在单一问题上的解决方案十分成熟, 例如磁盘异构问题、模型老化问题和小样本问题。然而, 由于这些问题经常同时存在, 只能处理其中一个问题的模型在实际预测中存在偏差。目前针对不同问题的解决方案经常相互冲突, 然而现有磁盘故障预测方法通常简单地融合各种模型, 缺乏在面对多个问题时对训练数据准备和学习模式的讨论。为此, 提出一种多属性数据划分方法 (MDP), 来探索针对多个问题的训练数据准备。引入与模型无关的元学习算法 (MAML), 对被划分的多个数据子集进行多任务学习。基于这些改进, 提出一种名为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。此外, MDP-MAML能够适应新出现的问题并进行学习和预测。在两个实际数据中心的数据集上, 与最先进方法相比, MDP-MAML将曲线下面积 (AUC) 从0.85提升至0.89, 将误检率 (FDR) 从0.85提升至0.91, 将误报率 (FAR) 从4.88%降低至2.85%。

**关键词:** 存储系统可靠性; 磁盘故障预测; 自我监测、分析及报告技术 (SMART); 机器学习

<https://doi.org/10.1631/FITEE.2200488>