

分类号_____

学校代码 10487

学号 M201572736

密级_____

华中科技大学

硕士学位论文

基于在线学习的磁盘故障预测技术

学位申请人：熊壮

学 科 专 业：计算机软件与理论

指 导 教 师：胡侃 副教授

答 辩 日 期：2018 年 5 月 30 日

**A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Engineering**

**Disk Failure Prediction in Data Centers
via Online Learning**

Candidate : Zhuang Xiong

Major : Computer Software and Theory

Supervisor : Prof. Kan Hu

Huazhong University of Science & Technology

Wuhan, Hubei 430074, P.R.China

May, 2015

独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：熊壮

日期：2018年5月28日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权华中科技大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本论文属于 ☐ 保密，在_____年解密后适用本授权书。
☒ 不保密。

（请在以上方框内打“√”）

学位论文作者签名：熊壮

日期：2018年5月28日

指导教师签名：胡悦

日期：2018年5月28日

摘要

随着数据中心中存储系统规模的极速增长，原本偶然发生的磁盘故障逐渐变为常态。为提高存储系统可靠性，国内外研究人员基于统计和机器学习的方法使用 SMART (Self-Monitoring, Analysis and Reporting Technology) 特征建立了许多磁盘故障预测模型，并取得了较好的预测性能。然而，这些工作大都是在离线模式下训练模型，没有考虑磁盘 SMART 属性分布规律随时间变化的影响，无法很好地拟合未来的数据模式，使模型在长期实际应用中预测性能会逐渐衰退，出现“模型老化”现象。

基于在线学习方法，设计实现的具有自适应性的在线故障预测系统，能避免预测性能的衰退问题，提高故障预测模型的实用性。运用在线学习方法对磁盘故障进行预测会面临以下挑战：1) 如何在线地对陆续收集的 SMART 样本进行标记？2) 如何克服正负样本不平衡对预测模型性能的影响？针对前者，提出了样本自动在线标记方法。该方法根据磁盘故障状态，实时地对 SMART 样本进行标记，作为训练数据输入到在线学习模型进行模型更新。其次，对 online bagging 方法进行改进，通过使用两个不同参数的泊松分布函数分别作用在实时的正负样本之上，使得负样本相对正样本具有更小的概率被选中而实际地用于模型更新。考虑到在线随机森林 (Online Random Forests, ORFs) 具有可并行性、低内存需求以及更好的预测性能等优点，在线磁盘故障预测系统基于 ORF 实现。

实验表明基于 ORF 的在线预测模型能够快速逼近离线随机森林 (Random Forest, RF)，并在低误报率 (False Alarm Rates, FARs) 前提下实现稳定于 93-99% 间故障检出率 (Failure Detection Rates, FDRs)。相较于进行离线更新的 RF 模型，ORF 模型能够在实现近似相同的 FDR 时维持更低的 FAR，并且无需每隔一段时间重新训练一个新的模型。因此，基于 ORF 的在线学习方法适用于实际的长期使用。

关键词：硬盘故障预测；在线学习；SMART；存储系统可靠性

Abstract

The rapid expansion of storage systems in data centers makes the originally accidental component failure become the norm. Disks are among the most frequently failed components. Based on SMART (Self-Monitoring, Analysis and Reporting Technology) attributes, many researchers derive disk failure prediction models using machine learning techniques and have achieved noticeable improvements. However, the majorities rely on offline training and fail to adapt to the future pattern of data, since the underlying distribution of SMART attributes changes over time. Consequently, the existing offline models suffer from the model aging problem and lead to declining performance in practice.

To address this problem, a novel way for disk failure prediction using the online learning method is presented. There are two main challenges one may encounter when training disk failure prediction models in online mode: 1) How to label the sequentially gathered samples on-the-fly? 2) How to overcome the impact of data imbalance on prediction performance? An automatic online label method and an improved online bagging method are respectively proposed to solve these challenges. Based on Online Random Forests (ORFs), an adaptive learning model is actually built to predict disk failure. The ORF-based model can automatically evolve with sequentially arrival data on-the-fly and thus is highly adaptive to dynamic distribution of SMART data. Moreover, it has favourable advantages against the offline counterparts in terms of parallelizability, low memory requirements and superior prediction performance.

Experiments on real-world datasets show that the ORF model converges rapidly to the offline random forests and achieves stable failure detection rates (FDRs) of 93-99% with low false alarm rates (FARs). Compared with offline RF models updated with offline updating strategies, the ORF-based models can maintain reasonably lower FARs while achieve comparable FDRs with these periodically updated RF models, however, without the need of model retraining after the initial deployment. Furthermore, we demonstrate the ability of our approach on maintaining stable prediction performance for the long-term usage in data centers.

Keywords: Disk failure prediction; Online learning; SMART; Storage system reliability

目 录

摘 要.....	I
ABSTRACT.....	II
1 绪论	
1.1 研究背景与意义	(1)
1.2 国内外研究现状	(3)
1.3 存在的问题与研究内容	(7)
1.4 论文的组织结构	(10)
2 在线磁盘故障预测关键技术	
2.1 在线随机森林算法	(12)
2.2 基于 ORF 的在线故障预测	(13)
2.3 本章小结	(19)
3 在线磁盘故障预测系统设计	
3.1 系统总体架构设计	(20)
3.2 子系统功能模块设计	(21)
3.3 本章小结	(23)
4 性能测试与评估	
4.1 测试数据	(24)
4.2 数据预处理	(26)
4.3 性能评估指标	(27)
4.4 ORF 模型性能评估	(28)
4.5 模拟实际长期使用	(34)
4.6 本章小结	(37)
5 总结及展望.....	(39)
致 谢.....	(41)
参考文献.....	(43)
附录 1 攻读硕士学位期间发表的论文.....	(47)
附录 2 攻读硕士学位期间发明的专利.....	(48)

附录 3 攻读硕士学位期间参与的科研项目..... (49)

1 绪论

1.1 研究背景与意义

数据是数据中心的核心，数据可用性是保障数据中心正常运行的关键。存储系统负责保存数据，提供数据访问的接口，是数据中心的主要组成部分之一。存储系统可靠性一直是研究人员与工业界关注的焦点。随着云计算和互联网服务的高速发展，由大量 IT 资源高度聚集而成的数据中心的存储系统规模变得越来越大，这使得原本偶发的硬件故障逐渐变为常态^[1]。而在所有硬件故障中，磁盘故障的占比远远高于其他硬件^[2]，其影响亦最为严重，会造成服务宕机甚至数据丢失等灾难性影响^[3]。作为存储系统的核心组成部分，磁盘能否提供稳定可靠的数据访问能力，直接影响整个存储系统的可靠性^[4]。因此，提高磁盘可靠性成为保障存储系统高效、稳定的关键。

磁盘是磁、电和机械的混合体，其固有结构决定了磁盘本身无法具有很高的可靠性^[5]。在现今的数据中心中，存储系统往往包含数十万甚至数百万块磁盘。当数据中心磁盘数量达到如此规模时，即便工艺的提升使得单个磁盘具有较低的故障发生率，但由于基数庞大，磁盘故障仍会大量发生。图 1-1 显示了 Schroeder 等人^[1]对三个数据中心中硬件故障发生的相对频率的统计结果。其中 HPC1 是一个包含 765 个结点的高性能计算集群，而 COM1 和 COM2 是两个互联网服务提供商。研究结果表明，磁盘故障的数量往往远高于其他硬件设备，比如 CPU、内存等，磁盘故障平均占比可以达到 30%，在某些数据中心中甚至会造成 50%左右的硬件替换。

对 BackBlaze 数据中心磁盘故障数据¹的研究表明，磁盘发生故障的几率还与磁盘生产商的生产工艺密切相关。具体而言，不同的磁盘生产商、不同生产批次的产品，其发生故障的几率存在很大的差异，即存在一些批次的磁盘，由于生产工艺的原因，磁盘可靠性较差，而导致故障发生频繁。另一方面，为了满足用户对高存储密度的需求，磁盘厂商的产品容量在不断扩大。伴随着磁盘容量的扩充，每块磁盘的扇区数量也随之增长，这使得磁盘发生扇区故障^[6]和数据损坏^[7]的几率大大增加。正如来自 Google 的数据中心研发人员所提出的，为了满足用户对高性能低成本的追

¹ The backblaze hard drive data and stats. <https://www.backblaze.com/b2/hard-drive-test-data.html>.

求，若要继续以用户可接受的较低价格生产磁盘，那么数据中心必须能够容忍磁盘具有较之前更高的故障率，而处理磁盘故障的任务则应该交由存储系统的更高层次来完成^[8]。

HPC1		COM1		COM2	
Component	%	Component	%	Component	%
Hard drive	30.6	Power supply	34.8	Hard drive	49.1
Memory	28.5	Memory	20.1	Motherboard	23.4
Misc/Unk	14.4	Hard drive	18.1	Power supply	10.1
CPU	12.4	Case	11.4	RAID card	4.1
PCI motherboard	4.9	Fan	8.0	Memory	3.4
Controller	2.9	CPU	2.0	SCSI cable	2.2
QSW	1.7	SCSI Board	0.6	Fan	2.2
Power supply	1.6	NIC Card	1.2	CPU	2.2
MLB	1.0	LV Power Board	0.6	CD-ROM	0.6
SCSI BP	0.3	CPU heatsink	0.6	Raid Controller	0.6

图 1-1 数据中心硬件故障的相对频率

数据中心的实际运营表明，磁盘故障相较于其他硬件故障带来的损失往往更加严重^[9]。一方面，磁盘故障很可能会造成数据的丢失，而在许多业务场景下，譬如银行、科研机构等，数据是极其敏感的，这些敏感数据的丢失无疑会造成灾难性的后果；另一方面，磁盘故障发生后若由于一些原因（大量磁盘集中故障，备份磁盘不足）而无法得到及时的处理，还会导致集群服务宕机，从而造成正在运行中的作业中断，严重影响作业的稳定运行时间。基于美国 63 个数据中心的一项研究²表明，数据中心发生服务宕机时的经济损失已从 2010 年的 5600 美元/分钟增加到了 2016 年的 8,851 美元/分钟。因此，为了保证数据中心能够提供稳定、可靠的服务，减小数据中心的拥有成本^[10]（包括运营成本、服务宕机损失和硬件故障成本等），提高存储系统可靠性变得至关重要。

传统的提高存储系统可靠性的方法主要是冗余磁盘阵列^[11] (Redundant Arrays of Inexpensive Disks, RAID) 技术。该技术通过数据冗余，可以允许单个或多个磁盘发生故障，并使用数据编码的方式对故障盘上的数据进行重构。在磁盘层面上，可以通过数据冗余来提高数据可靠性；在系统层面上，则可以通过数据副本等技术提高存储系统可靠性。Google 数据中心的文件系统 GFS^[12]通过采用随机副本技术来提高数据修复能力。Renesse 等人^[13]研究了链式副本技术和伪随机的分布方法用于提高数

² Data center downtime costs. <http://www.emerson.com/en-us/News/Pages/Net-Power-Study-Data-Center.aspx>.

据可靠性。上述技术都属于被动容错，其出发点在于“容忍”磁盘故障，即在故障发生后，采取相应的方式恢复或重构故障盘数据，使得存储系统仍然能够提供数据访问服务。然而这类方法会影响磁盘读写性能，增加故障发生率，降低存储系统的整体性能。

主动容错技术则是根据磁盘的历史表现，在故障真实发生前对潜在的磁盘故障进行预测，并依据预测结果对存在较大风险的磁盘进行数据迁移^[14]或者换盘等处理，从而有效避免磁盘故障的发生。SMART^[15]就是一种典型的主动容错技术，几乎所有的磁盘厂商都在其产品中应用了该技术。该技术实时监测磁盘内与健康状况有关的技术监测值，比如磁盘温度、寻道出错率、重映射扇区数等信息，当某一个监测值超过了厂商预先设定的阈值，就发出预警信号，这就是基于 SMART 的原生阈值方法。这种方法简单易行，但为了保证低误报率 (False Alarm Rate, FAR)，磁盘厂商对阈值的设置很严谨，导致故障检出率 (Failure Detection Rate, FDR) 只有 3%-10%^[16]。

为了提高磁盘故障预测的预测精度，研究人员基于大量磁盘 SMART 数据和磁盘故障数据，采用统计或者机器学习的方法建立磁盘故障预测模型。这些故障预测模型能够在磁盘故障发生之前对其进行预测，用户则根据预测结果对存在较大风险的磁盘进行数据转移或者更换磁盘等处理，从而避免磁盘故障真实发生，达到增强存储系统可靠性的目的。

1.2 国内外研究现状

SMART 技术起源于 IBM 在 1992 年提出的可预测故障分析技术 (Predictive Failure Analysis, PFA)，该技术周期性地对各个磁盘属性进行检测，当任意一个属性的数值超过其预定义的阈值时，其会向用户发送预警信息。PFA 技术得到了工业界的普遍接受，并于 1994 年进一步发展为 SMART 标准，用于 ATA 和 SCSI 磁盘的可靠性预测。SMART 的发展经历了以下三个阶段^[18]：

- (1) 初始阶段的 SMART 仅对在线活动进行监控，从而提供对故障的预测；
- (2) 后续阶段的 SMART 增加了自动离线扫描用于对额外磁盘操作进行监控；
- (3) 最后阶段的 SMART 在监控磁盘运行活动的同时还增加了故障预防机制，自动地尝试对扇区错误进行修复。

目前几乎所有磁盘厂商均支持 SMART 技术，作为行业标准，SMART 规定了磁盘生产商应遵循的准则。SMART 在磁盘空闲时通过“离线数据收集”对所有扇区进

行测试，以确认磁盘是否健康。SMART 信息保留在磁盘的系统保留区(Service Area)内，这个区域一般位于磁盘最前面几十个物理磁道，由厂商写入相关内部管理程序。磁盘厂商常用的故障预警方法是阈值算法，该方法简单易行，但是预测精度太低。

为了提升预测系统的预测精度，研究人员基于各种统计学和机器学习方法提出了许多磁盘故障预测技术。这些磁盘故障预测技术大都使用磁盘 SMART 属性构建预测系统，且具有相似的流程：先根据磁盘状态数据对磁盘 SMART 样本进行标记，故障磁盘的样本标记为正样本 (positive sample)，而正常磁盘的样本则标记为负样本 (negative sample)，进而组织成包括特征向量和分类标签的矢量，然后利用统计学或机器学习方法训练得到一个分类模型，最后根据磁盘的实时 SMART 数据进行故障预测。各故障预测技术的比较见表 1-1。

表 1-1 各故障预测技术比较

方法	相关研究	特点
阈值法	[18]	需要人为设定的模型参数较多，且参数对模型性能影响非常大
贝叶斯分类	[19][22][23]	FDR 不高，需要样本特征不相关的假设
秩和检验	[20]	经典统计学方法，FDR 仅 60%
SVM	[21][31]	FDR 一般，FAR 很低，模型训练时间长，参数较多
HMM	[24]	考虑磁盘样本时序性，但模型训练时间长，FDR 不高
马氏距离	[25][26][27]	计算效率很高，但可解释性差，FDR 仅 68%
BP NN	[31]	FDR 较高，但 FAR 也较高，模型可理解性差
决策树	[32][35]	预测性能好，但模型易产生过拟合，FDR 可达 95.5%
规则	[36]	可理解性好，故障检测率与规则数量成正比，误判率较高

Hamerly 等人^[19]基于贝叶斯方法提出了两种不同的磁盘故障预测技术：朴素贝叶斯子模型混合 (Naive Bayes submodels, NBEM) 和朴素贝叶斯分类器 (supervised naive Bayes classifier)。前者使用期望最大化算法 (Expectation-maximization, EM) 训练数据；后者的优势在于算法简单且易于实现。两个算法都在来自昆腾公司的数据集上进行测试，该数据集中包含 1927 个良好磁盘和 9 个故障磁盘，在误报率约为 1% 时实现的故障检出率分别为 35-40%、55%。

Hughes 等人^[20]基于统计理论提出了两个改进的预测算法。他们对数据进行统计分析时发现，许多 SMART 属性都是非参数分布的。基于这个发现，他们使用秩和检验 (rank-sum test) 对磁盘故障进行预测，以降低误报率的同时提高故障预测的精

度。在 3744 个磁盘（其中有 36 个故障盘）上进行的实验表明，他们的算法在 0.5% 误报率条件下，可以实现 60% 的故障检出率。

Murray 等人^[21, 22]比较了不同的机器学习方法在磁盘故障预测领域的表现，他们将磁盘故障预测问题抽象为在非参数化分布且充满噪音的数据中对小概率事件进行检测。其中，文献[21]对支持向量机 (Support Vector Machine, SVM)，非监督聚类 (unsupervised clustering) 和两个非参数化统计方法 (rank-sum 和 reverse arrangement) 进行了比较。实验中使用的数据集来自于 369 个磁盘，这些磁盘的磁盘模型 (disk model) 均相同，且其中故障磁盘与正常磁盘各占一半。实验结果表明，SVM 和 rank-sum 的预测性能要远远优于磁盘厂商所使用的基于阈值的方法，且误报率也更低。他们发现，在使用确定的某小部分 SMART 属性时，rank-sum test 在 FAR 为 0.5% 时 FDR 为 33.2%，性能要优于 SVM；而使用全部所选的 25 个特征时，SVM 性能最好，其 FDR 为 50.6% 而 FAR 为 0%。文献[22]基于多实例学习框架和贝叶斯分类器提出一个新的算法 mi-NB (multiple-instance Naive Bayes)，其能够实现较低的误报率和较好的性能。

Tan^[23]和 Zhao^[24]分别提出树扩张朴素贝叶斯方法 (Tree Augmented Naive Bayesian, TAN) 和隐式马尔可夫模型 (Hidden Markov Model, HMM) 来进一步提高磁盘故障预测系统的故障检出率。其中，文献[23]的结果表明，当 TAN 的 FAR 为 3% 时，其 FDR 可以达到 80%；但是当约束 TAN 实现零误报时，其 FDR 只能达到 20-30%。文献[24]使用与文献[21]相同的数据集，他们的结果表明，当 FAR 为 0% 时，HMM 方法能实现 52% 的故障检出率，但是当误报率增加时，并不能获得比 TAN 方法更好的故障检出率。

Wang 等人^[25]在 2011 年提出基于马氏距离 (Mahalanobis Distance, MD) 的磁盘检测方法。使用跟[21]中相同的数据集，他们证实由失败模式、机制和效果分析机制 (Failure Modes, Mechanisms and Effects Analysis, FMMEA) 选择出来的优先属性比全部属性得到的预测性能要好。在他们随后的研究^[26]中，最小冗余最大相关性 (the minimum redundancy maximum relevance, mRMR) 理论被用来从由 FMMEA 选择出来的优先属性中去除多余的属性。接着，他们使用良好磁盘的关键属性建立了一个的基准马氏空间，这个模型能够预测出 67% 的故障磁盘且其 FAR 为 0，其中有 56% 的故障磁盘能被提前 20 个小时预测出来。

Wang 等人^[27]在后续工作中进一步提出了一个用于磁盘故障预测的两步统计方法。该方法包括两个步骤：异常检测和故障预测。在异常检测阶段，该算法使用马氏距离将 SMART 变量转变为一维索引，再使用 Box-cox 变换将其变换为高斯变量，然后通过定义阈值，检测出磁盘异常。在故障预测阶段，算法通过基于滑动窗口的通用似然比来对磁盘异常的变化情况进行跟踪，在某一时间段内，如果发现磁盘异常频繁发生，则意味着磁盘很可能发生故障。实验结果表明，该方法在零误报时可以实现 68% 的故障检出率。除了预测性能上的提升，该方法通过将多元的 SMART 数据转化为单变量数据，极大地提高了预测算法的计算效率；通过对关键特性子集使用基于规则的故障检测算法，增强了预测的准确性。

在国内方面，华中科技大学刘景宁等研究人员^[28]对 RAID 可靠性进行了研究，提出了一种基于 SMART 的 RAID 高可靠性方法。国防科学技术大学张超等研究人员^[29]提出了基于 SMART 信息的 T2US (Time, Temperature, Utilization and S.M.A.R.T) 算法预测磁盘故障，并实现了 52% 的预测准确率。他们在此基础上，还构建了多级 RAID 结构，这在一定程度上进一步提高了 RAID 的可靠性和性能。国防科学技术大学胡维等研究人员^[30]提出了一个支持智能预警和自动修复的高可靠磁盘阵列，他们使用提升决策树算法构建分类器，进一步提高故障预测性能。

以上这些模型的预测性能都没有很令人满意，一个可能的原因是因为上述模型使用的数据规模相对偏小，无法覆盖足够充分的 SMART 信息以用于建立有效的预测模型。数篇文献训练和测试模型所使用的数据集都是由 Murray 等人^[21]提供的数据集，然而该数据集仅包含 369 个磁盘，其中故障磁盘与正常磁盘各占一半，这明显不符合现实数据中心的情况。此外，该数据集是在 2003 年以前采集的，其 SMART 属性含义与编码方式不符合当前的 SMART 标准。毫无疑问，这些因素削弱了模型的实用性。

为了克服数据集相对太小、数据集 SMART 属性含义与编码不符合当前的 SMART 标准的问题，Zhu 等人^[31]使用来自现实数据中心的包含 23,395 个磁盘的数据集进行模型的训练与测试。他们探索反向传播人工神经网络模 (Backpropagation artificial neural networks, BP ANN) 的预测性能。实验表明，BP ANN 模型能实现很好的预测性能，其在 FAR 在合理范围内时能够实现高达 95% 的 FDR。

上述诸多统计和机器学习的模型，尽管它实现了较好的预测性能，但在应用于实际的数据中心时表现不理想，主要表现在其稳定性不好，可解释性较差，预测模型

难以调整,并且这些模型不能准确地描述硬盘性能的逐渐恶化。为了解决这些问题, Li 等人^[32]基于分类和回归树 (Classification and Regression Trees, CART) 理论建立了新的硬盘故障预测模型,其预测性能以及稳定性、可解释性较 BP ANN 模型都要好。实验结果表明,CT 模型在包含 25,792 个磁盘的真实数据集上可以达到如下的预测性能: $FDR > 95\%$ 且 $FAR < 0.1\%$ 。实验中用不同的驱动集和不同的模型更新策略来训练预测模型,因此其稳定性很高。

之前提及的所有理论包括 CT 模型其实质都是二元分类器,这些模型对于硬盘的评估都只能得到一个好/坏标签,而不能有效的描述硬盘的健康状况,然而硬盘的恶化是一个逐渐的过程,仅根据标签好坏就更换硬盘使成本较大。基于 RT 理论的硬盘健康指数模型则可以得到各个驱动的健康评估而不仅仅只是得到好/坏标签,因此其能够依据硬盘健康指数的排序来处理模型发出的警告,即可以优先地处理那些更加趋向损坏的硬盘,可以有效降低成本并且模型可解释性较好。

在接下来的研究中, Xu 等人^[33]和 Li 等人^[34]也将磁盘故障预测问题作为回归问题进行处理,他们分别使用递归神经网络 (Recurrent Neural Networks, RNN) 和梯度提升回归树 (Gradient Boosted Regression Trees, GBRTs) 训练剩余寿命预测模型 (residual life prediction models),用于磁盘健康程度的预测。在来自于真实数据中心的大型数据集上,RNN 模型在正样本上的 ACC (accuracy of residual life level assessment) 达到了 40-60%^[35]。

1.3 存在的问题与研究内容

1.3.1 存在的问题

之前工作都是基于一段固定的时间内采集到的样本数据在离线模式下进行模型的训练与评估,这些模型被假设为在创建之后就不会改变^[32]。尽管这些模型能很好的拟合该段时间内所采集样本数据的特征,而在使用初期具有理想的表现,但是,在实际的长期使用时,模型的预测性能并不稳定而是逐渐变差,这个问题被称为“模型老化 (model aging)”。存在模型老化的原因是,数据中心新采集的数据会使得部分 SMART 属性的分布规律发生显著的变化,这就导致先前建立的模型无法很好的适应当前数据的特征,从而使得预测模型逐渐失去效果。

分布规律会发生动态变化的 SMART 属性往往都是累计值属性 (cumulative attributes),而这些属性正好是对磁盘故障的预测起关键作用的属性。以属性 POH

(Power-on Hours) 作为累计值属性的代表, 图 1-2 描绘了一个数据集中该属性随时间的分布规律的变化, 该图分别表示了 2014 年和 2015 年间采集的样本数据的该 SMART 属性的分布规律。从图中可以看出, 随着时间推移, 数据集累计值属性的分布会发生较大的变化。使用 2014 年间采集的数据训练预测模型, 其学习得到的 POH 属性的阈值为 87 (即图 1-2 中的 old threshold), 也就是说, 若属性 POH 的值低于 87, 则会触发报警, 反之说明磁盘暂时健康。而 2015 年间由于属性分布的变化, 其阈值也随之改变为 70 (即图中 new threshold)。若不对模型进行更新, 使用 2014 年的数据训练的模型对 2015 年间的数据进行预测, 则会导致 POH 值在 70-87 之间的所有样本数据都错误的触发报警, 而带来大量的误报。

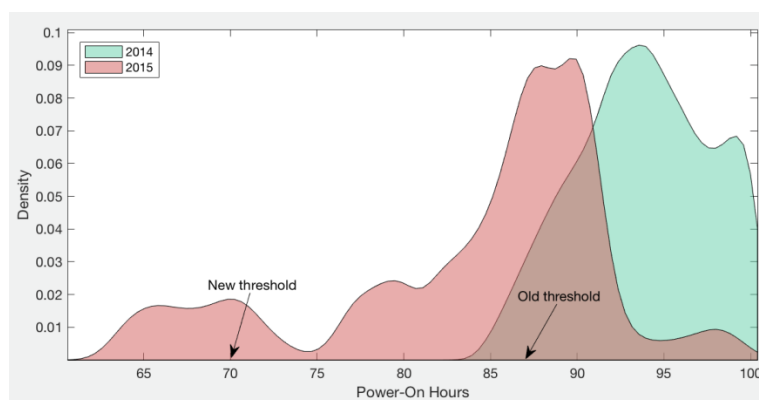


图 1-2 POH 属性的分布规律随时间变化

为了解决模型老化的问题, 因为当前提出的磁盘故障预测模型全部是离线模型, 现存的解决方案都是对模型进行离线的更新。Li 等人^[32]提出了两种离线模型更新策略: *replacing* 策略和 *cumulation* 策略。两个策略都是周期性地对模型进行更新, 在他们的情况下每周更新一次。*replacing* 策略更新模型时仅使用最近一周内采集的数据, 而 *cumulation* 策略则是使用从一开始采集到最近这一周的所有数据进行模型更新。更新之后的模型用于对下一周采集的样本数据进行预测。尽管这两个策略都能使模型预测精度得以维持在一个较好的水平, 即能有效解决模型老化问题, 但是这些策略都无法避免每隔一段时间就需要结合新采集的数据重新训练一个新模型替换之前的模型。因此, 离线地训练或者更新模型都无法使模型自适应数据的动态变化。

一种能从根本上的解决模型老化的方案是, 直接部署能够自动进化的学习模型, 这意味着模型可以自动地对陆续到来的数据进行学习, 而使得模型始终适应当前的数据特征。在线学习^[37] (online learning) 就是这样一种方案。在线学习相较于之前的离线方法具有许多优势, 例如: 内存需求要小很多, 因为离线模式下需要同时学习

的大量数据在在线时被分散到每天进行学习；训练效率往往较高，训练及预测时间往往更短；此外，在线学习还能够实现实时的故障预测。

1.3.2 主要研究内容

将在线学习方法运用到磁盘故障预测时会面临两个挑战，本课题针对这些挑战分别提出了相应的解决方案。

第一个挑战是当模型在在线模式下进行学习时如何对 SMART 样本进行标记。离线模式下，由于所有的训练数据都已采集完成，并且数据集中各个磁盘故障与否也是事先告知的，因此离线模式下对 SMART 样本标记很简单，只需要根据各个磁盘的故障信息对该磁盘的 SMART 样本进行标记即可。而在线模式下，模型是在线地结合新到来的训练数据进行学习的，与此同时该模型还用于对新采集的样本进行预测。在线学习时，新采集的 SMART 样本是无法进行标记的，因为磁盘当前的健康状态不确定，也就是说无法预知当前数据中心各磁盘是否即将故障。具体而言，数据中心中许多磁盘虽然仍在运行但其健康状态可能已较差，即很可能在近期发生故障，然而对于这些不太健康的磁盘，他们报告的 SMART 样本也仅能在磁盘真实发生故障之后才被标记为 positive，这部分样本也不能直接标记为 negative，因为磁盘不是处于健康状态的。因此，对于这些不太健康的磁盘，其在真实故障之前报告的 SMART 样本是无法进行标记的。针对这个问题，本文提出了一个自动的在线标记 (online label) 方法，它会为每一块磁盘创建一个定长的队列用于保存该磁盘最新 7 天时间内采集的样本。暂存于队列中的样本是未标记的，而溢出队列的样本会全部被标记为 negative。仅当一块磁盘实际发生故障后，才将该磁盘的样本队列中的所有样本标记为 positive。一旦样本被标记之后，其就作为新的训练数据输入在线学习模型用于模型的更新。

第二个挑战是正负样本的严重不平衡。由于磁盘故障是一个小概率事件，故障磁盘的数量往往要远小于健康磁盘，这就导致故障样本（正样本）要远少于健康样本（负样本）。实验表明负样本的数量可能是正样本的几百上千倍。离线模式下，可以通过对负样本进行采样，以得到跟正样本数量相当数量的负样本。而在线时，由于训练数据是逐渐到来的，因此样本采样无法进行。为解决样本不平衡的问题，本文对 Oza 等人^[38]提出的 online bagging 进行改进，得到了一个可行的解决方案。该方案通过使用参数不同的两个泊松分布函数分别作用于陆续到来的正样本和负样本，

使得负样本被选中用于更新的几率远小于正样本。实验表明，当选取合适的柏松参数时，该方法能很好的解决样本不平衡对模型预测性能带来的影响。

此外，本文还基于在线随机森林^[39] (Online Random Forests, ORFs) 具体实现了一个用于磁盘故障预测的在线学习模型。相较于之前的工作所使用的机器学习方法，ORF 算法具有诸多明显的优势，比如：其训练和预测过程能够并行化，因为森林中的每一颗树都是独立于其他树而创建的；预测模型具有高度可解释性，因此可以用于揭示造成磁盘故障的真正原因从而帮助提升存储系统可靠性^[36]。更重要的是，ORF 算法能够结合新到来的 SMART 数据在线地生成随机树 (random trees)。如果某一些树的 OOB^[40] (Out-Of-Bag-Error) 太大，则说明这些树已太过陈旧而无法适应当前数据的模式，ORF 算法会丢弃这些树，并生成新的树作为替换。因此，本文中基于 ORF 的预测模型对于 SMART 属性分布规律的变化具有很好的自适应性，并且其对于标记噪音也具有很好的鲁棒性^[41]，这使得本文提出的在线学习方法很适合在实际数据中心中的长期使用。

1.4 论文的组织结构

本文主要研究如何将在线学习方法运用到磁盘故障的预测，以解决离线模型所存在的模型老化问题。论文分为五个章节进行讲解，具体的组织结构如下：

第一章首先介绍了课题的研究背景，说明了磁盘故障在当今数据中心中发生频繁，在所有硬件故障中占比最大，且影响最为恶劣，因此对于提升数据中心服务质量，磁盘故障的预判至关重要。接着概述了国内外有关磁盘故障预测技术的研究现状，分析了之前的工作中所使用的各种统计学或机器学习方法的优劣。然后提出了这些工作存在的问题，即离线模型的模型老化问题，并围绕该问题介绍了本课题的主要研究内容。最后说明了论文的组织结构。

第二章为论文的核心内容，首先简单介绍了在线随机森林算法以及该算法的优势，接着具体讲解了将在线算法运用于磁盘故障预测时会面临的两个挑战：在线模式下的样本标记问题和磁盘故障正负样本不平衡问题。针对这些挑战，本章提出了可行的解决方案，并给出了优化后的 ORF 算法和基于 ORF 的在线学习算法。

第三章讲解了在线磁盘故障预测系统设计方案，将系统分为数据预处理模块、模型训练模块、样本在线标记模块、模型在线更新模块和磁盘结果预测模块等五个模块分别进行讲解。

第四章对本文提出的在线学习方法进行了详细的评估。首先介绍了实验评估中所使用的数据集的相关统计信息，接着讲解了对数据进行的预处理操作，即关键特征选取。然后将 ORF 模型与离线的 RF 模型、SVM 模型和 DT (Decision Tree) 模型进行对比分析，说明了本文提出的 ORF 模型的预测性能。最后模拟了在线模型和离线 RF 模型在实际数据中心中的长期使用，说明了本文提出的在线模型能够很好的克服模型老化的问题。

第五章对全文进行了总结，同时对未来的研究工作进行展望。

最后是致谢和参考文献说明。

2 在线磁盘故障预测关键技术

相较于传统的离线磁盘故障预测，在线磁盘故障预测会面临诸多挑战，比如在线学习时如何解决样本标记的问题，以及在线时如何解决样本不平衡的问题。本章的核心内容就是解决将在线学习方法应用到磁盘故障预测时所面临的挑战，并给出适用于磁盘故障预测的基于 ORF 的在线学习技术。在内容安排上，本章首先介绍 ORF 模型的相关知识，然后具体讲解在线故障预测会面临的挑战以及本课题提出的解决方案，同时给出优化后的 ORF 算法和基于 ORF 的在线学习算法。

2.1 在线随机森林算法

在线随机森林^[39] (Online Random Forests, ORFs) 常用于分类或者回归问题，是随机森林^[41]以及在线机器学习^[37]的结合。给定一个由两个类组成的训练集，该集成方法在训练阶段中会构造许多颗决策树，在预测阶段则根据这些树的投票结果对样本进行预测。

为了能够在线的进行学习，ORF 算法借鉴了 online bagging 技术^[38]和在线树生成技术^[39]。为了实现与离线 bagging 相同的效果，即每个训练样本被森林中的每棵树随机选取，ORFs 使用 Oza 等人提出的 online bagging 方法，该方法应用泊松分布函数来限制顺序到达的数据是否被选取。Oza 等证明了这种方法的收敛性。

在 ORF 的每棵随机树的生长过程中，每个决策节点都会创建一组随机测试，并根据一些不纯度测量手段从中选择最佳的测试作为结点的划分函数。就硬盘故障预测而言，每个随机测试的形式为 $SMART_i > \theta$ ，其中 $SMART_i$ 表示 ID 编号为 i 的 SMART 属性，而 θ 是该属性的一个阈值，用于决定决策结点中 SMART 样本的左右划分。为了评估包含 N 个随机测试的集合 $S = \{s_1, \dots, s_N\}$ 中每个测试 s 的质量，该算法适用 *Gini Impurity* 作为不纯度测量标准。决策结点 D 的 *Gini Impurity* 计算公式如公式(2-1)所示：

$$G(D) = p_0(1 - p_0) + p_1(1 - p_1) \quad (2-1)$$

其中 p_0 和 p_1 由公式 $p_0 + p_1 = 1$ 限制，分别表示结点 D 中两个类的占比。 $G(D)$ 取值于 0 到 0.5 之间，且结点 D 的 *Gini Impurity* 越大，说明该结点的纯度越低。举例说明，若 $G(D) = 0$ ，则 $p_1 = 1, p_0 = 0$ 或者 $p_1 = 0, p_0 = 1$ ，此时结点 D 中所有样本都属于同

一类，纯度最高；若 $G(D) = 0.5$ ，则 $p_1 = 0.5$ ， $p_0 = 0.5$ ，此时结点 D 中两类样本各占一半，纯度最低。对于测试集合中的一个测试 $s \in S$ ，假设结点 D 根据这个测试被划分为左子结点 D_{ls} 和右子结点 D_{rs} ，则测试 s 的信息增益计算公式如公式(2-2)所示：

$$\Delta(D, s) = G(D) - \frac{|D_{ls}|}{|D|} G(D_{ls}) - \frac{|D_{rs}|}{|D|} G(D_{rs}) \quad (2-2)$$

其中 $|D|$ 表示结点 D 中包含的样本数目。结点划分的目标是减小不纯性，因此结点 D 划分得到的子结点越纯越好，也就是说子结点的不纯性度量越小越好。当一个结点满足划分条件的时候，算法先创建 N 条随机测试，然后分别计算按这些测试进行结点的划分时得到的信息增益，具有最大信息增益的测试就被选中该结点的划分函数，进而将该结点划分为两个子结点。

模型在线学习时，训练数据是逐渐到来的，ORF 的每棵树的每个结点都是在线的对落到该结点的 SMART 样本进行统计分析。算法使用两个超参数 *MinParentSize* 和 *MinGain* 来控制结点的分裂。参数 *MinParentSize*（标记为 α ）限制一个结点在分裂之前需要包含的最小样本条数，只有当结点中的样本条数不小于 α 时该结点才可能开始分裂，这一参数能够避免结点参数鲁棒性太差的统计计量。参数 *MinGain*（标记为 β ）限制了结点划分需要满足的最小信息增益，也就是说结点条数在满足不小于 α 后，还需要该结点创建的 N 条测试中至少有一条测试的信息增益不小于 β ，才能使用该测试进行结点的划分，这一参数用于限制结点划分的最小价值。因此，结点 D 进行划分需要满足的必要条件是： $|D| \geq \alpha$ and $\exists s \in S: \Delta G(D, s) \geq \beta$ 。

2.2 基于 ORF 的在线故障预测

以前的研究工作都是利用离线机器学习方法来训练磁盘故障预测模型，它们的正常功能依赖于整个训练数据都是可访问的这一假设，并且需要足够的有效信息来构建高性能模型。然而在实际中，训练数据通常是逐渐收集的，而不是提前给出的。这有可能导致初始阶段训练数据不够充分，从而成为模型性能的瓶颈。因此，部署支持在线学习的预测模型是更加明智的选择。更重要的是，在线学习支持利用顺序到达数据自动进行演化更新，因此可以对 SMART 数据分布规律的动态变化保持高度的适应性，从而消除模型老化问题。

本课题提出的在线学习技术基于 ORF 算法进行预测模型的训练，该算法继承了 RF 的所有优点，并且对于磁盘故障预测具有许多显著的优势。首先，ORF 的整体训练和测试过程可以很容易地进行并行化，因为森林中的每棵树都是独立于其他树进

行构建和测试的,这使得 ORF 的时间效率远高于梯度增强 (gradient boosting) 方法,例如 GBDT (Gradient Boosted Decision Trees, 梯度促进决策树)。其次, ORF 模型具有很高的解释性,这可以用于揭示磁盘故障的真正原因,并有助于提高存储系统的可靠性。此外,与 boosting 和其他集成方法相比, ORF 对于标签噪声 (label noise) 也具有更强的鲁棒性。

2.2.1 样本不平衡解决方案

磁盘 SMART 属性被用作构建预测模型的解释变量,而响应变量则用来表示磁盘在未来 7 天内是否会发生故障。训练出预测性能较好的预测模型的一个主要挑战是 SMART 训练数据的高度不平衡分布,因为健康磁盘占绝大多数, negative (健康磁盘) 样本要远远多于 positive (故障磁盘) 样本。值得注意的是,在原始的 ORF 算法中,所有的训练样本都是由一个相同的泊松分布函数来进行处理的,也就是说,原始 ORF 算法相同地对待训练样本中的正类或者负类。直接使用原始 ORF 算法进行预测模型的训练,很可能会导致预测模型出现较差的预测性能,因为分类算法通常是以最大化总体精度为模型优化的目标,这会使预测模型严重偏向于样本数量显著更大的类,在磁盘故障预测中,也就是会偏向 negative 类。

为了解决样本不平衡的问题,本课题对 online bagging 方法进行改进,提出了一种针对在线学习中样本不平衡的解决方案,并提出了两个超参数: λ_p 和 λ_n 。该解决方案分别使用两个不同的泊松分布函数对陆续到来的正负样本进行约束,以使得负样本具有比正样本更小的几率被选中并用于模型演化更新。假设随机森林中的第 t 颗树为 $f_t(x)$, 整个森林为 $\mathcal{F} = \{f_1, \dots, f_T\}$, 其中 T 表示森林中树的棵树。当一个训练样本 $\langle \mathbf{x}, y \rangle$ 到来时,森林中的每一棵树 $f \in \mathcal{F}$ 都将使用该样本重复进行 k 次更新。针对每一个到来的 positive 样本, k 由泊松分布函数 $\text{Poisson}(\lambda_p)$ 得来; 针对每一个到来的 negative 样本, k 则由泊松分布函数 $\text{Poisson}(\lambda_n)$ 得来。该方案可以使用以下公式(2-3)表示:

$$k(\langle \mathbf{x}, y \rangle) = \begin{cases} \text{Poisson}(\lambda_p), & y = 1 \\ \text{Poisson}(\lambda_n), & y = 0 \end{cases} \quad (2-3)$$

其中 $k(\langle \mathbf{x}, y \rangle)$ 表示样本 $\langle \mathbf{x}, y \rangle$ 的更新次数。超参数 λ_p 和 λ_n 通常设置为固定的数值,在接下来的研究过程中,参数 λ_p 被设定为 1, 这样针对每一个 positive 样本,其被用于模型更新的次数 k 取得非零值的概率大约为 70%; 参数 λ_n 则通常被设定为远小于 1 的小数,比如 0.01, 这样针对每一个 negative 样本,其被用于模型更新的次数 k 取

得非零值的概率大约为 1%。这样的参数设置可以使得陆续到来的 negative 样本被选中用于模型更新的几率远小于 positive 样本，从而使得实际用于模型更新 negative 样本数量跟 positive 样本数量相当，进而解决陆续到来的样本的不平衡问题。

对于作为比较的离线模型而言，模型的训练同样会面临样本不平衡的问题，为此本课题还针对离线模型训练提出了一个超参数：*NegSampleRatio* (标记为 λ)。给定一个训练数据集 D ， D_p 表示数据集 D 中 positive 样本组成的集合， D_n 表示数据集 D 中 negative 样本组成的集合。训练离线模型时，不再使用样本不平衡的原始训练数据集 D ，而是使用它的子集 $D_p + D_{nc}$ 作为实际的训练集。其中 D_{nc} 是 D_n 的一个子集，是从 D_n 中随机取样的结果，而参数 λ 就是用来控制这个取样容量。具体而言，从 D_n 中选取的 negative 样本数量是 D_p 中包含的 positive 样本数量的 λ 倍，可以用以下公式(2-4)表示：

$$\lambda = \frac{|D_{nc}|}{|D_p|} \quad (2-4)$$

参数 λ 通常设置为大于 1 的数。该参数对离线模型预测性能的影响可以参见本文 4.4 部分。

基于 ORF 的在线学习模型除了能够在线地结合陆续到来的 SMART 数据学习新的知识，还能够通过丢弃已经老化的树来抛掉过于陈旧的知识，这在 SMART 数据的分布规律会随时间发生变化时变得至关重要。对于森林中的每一棵树 $f \in \mathcal{F}$ ，仅当 k 通过 $\text{Poisson}(\lambda)$ 取值为正整数时，样本 $\langle \mathbf{x}, y \rangle$ 才会被用于模型的更新；而当 k 取零时，样本 $\langle \mathbf{x}, y \rangle$ 则会用于更新树 f 的 OOB (Out-Of-Bag-Error，袋外误差)。基于每一颗树的 OOB 和该树的年龄（用树包含的样本数量表示），ORF 算法会将树龄太大或者 OOB 太大的树丢弃掉，并且创建新的树作为替换。由于在森林中单颗树的作用相对是很小的，因此丢弃一颗树对于整个模型而言没有太大的损失，然而持续地创建能更好的拟合当前数据特征的新树用以替换预测误差太大的老化的树，可以使得预测模型对于训练数据分布规律的变化具有很好的自适应性。基于磁盘故障预测进行优化后的 ORF 如算法 1 所示。

算法 1 优化后的 ORF

输入： 陆续到来的训练样本 $\langle \mathbf{x}, y \rangle$

森林中树的数目 T

MinParentSize α , *MinGain* β

样本平衡参数 λ_p 和 λ_n

输出：用于磁盘故障预测的 ORF 模型

```
1: // For each one of T trees
2: for t from 1 to T do
3:   if y == 1 then
4:      $k \leftarrow \text{Poisson}(\lambda_p)$ 
5:   else
6:      $k \leftarrow \text{Poisson}(\lambda_n)$ 
7:   end if
8:   if k > 0 then
9:     // Update the tree  $f_t$  for k times
10:    for i from 1 to k do
11:       $j = \text{findLeaf}(x)$ .
12:       $\text{updateNode}(j, \langle x, y \rangle)$ .
13:      if  $|R_j| > \alpha$  and  $\exists s \in S : \Delta L(R_j, s) > \beta$  then
14:        // Find the best split from S
15:         $s_j = \arg \max_{s \in S} \Delta L(R_j, s)$ .
16:         $\text{createLeftChild}(p_{jls})$ 
17:         $\text{createRightChild}(p_{jrs})$ 
18:      end if
19:    end for
20:  else
21:    // Update OOB of tree  $f_t$ 
22:     $\text{OOBE}_t \leftarrow \text{UpdateOOBE}(\langle x, y \rangle)$ 
23:    // Estimate whether  $f_t$  is decayed
24:    if  $\text{OOBE}_t > \theta_{\text{OOBE}}$  and  $\text{AGE}_t > \theta_{\text{AGE}}$  then
25:      // Replace the outdated tree with new tree
26:       $f_t = \text{NewTree}()$ 
27:    end if
21: end if
22: end for
23: Output the forest F.
```

2.2.2 在线样本标记解决方案

将在线学习应用到磁盘故障预测会面临的另一个挑战是样本的标记问题。离线训练的时候，整个数据集都是给定的，并且数据集中各个磁盘的故障信息也是已知，因此离线模式下样本的标记只需根据所述磁盘的状态进行。具体而言，故障磁盘的

样本中在故障真实发生之前一段时间的样本被标记为 **negative**，而正常磁盘的样本则标记为 **positive**。在线模式下，训练样本是随时间逐渐到来的，在线模型也是使用这些逐渐到的样本进行演化学习，而如何对刚采集的 SMART 样本进行标记则成为一个不好解决的问题。数据中心中部分磁盘虽然仍在运行，但其健康状态可能已经较差，也就是说，这些亚健康状态的磁盘很可能在未来几天内发生故障，正是这些亚健康状态的磁盘所报告的 SMART 样本无法进行标记。如果将这些样本标记为 **negative**，但磁盘若在近期发生故障，这样的标记明显会成为标记噪音 (label noise)；这些样本也不能直接标记为 **positive**，因为只有当磁盘故障真实发生之后，其故障前一段时间内报告的样本才能标记为 **positive**。因此，本课题提出了一种在线标记方法，该方法会将所有磁盘最近一周内报告的样本暂存起来并不对其进行标记，直到采集了新样本或者磁盘故障真实发生。图 2-1 和图 2-2 表示了在线样本标记方法。

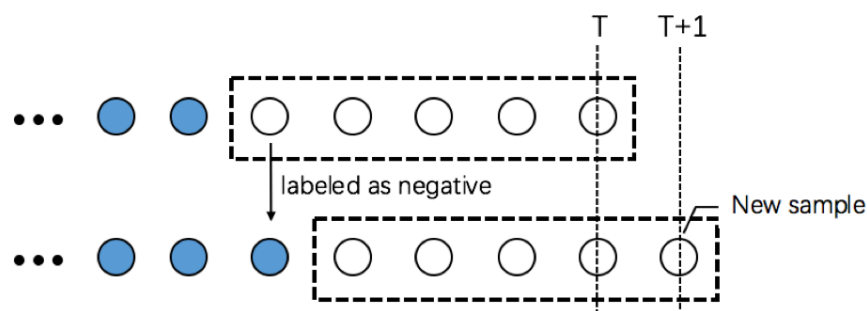


图 2-1 新样本到来时样本的标记

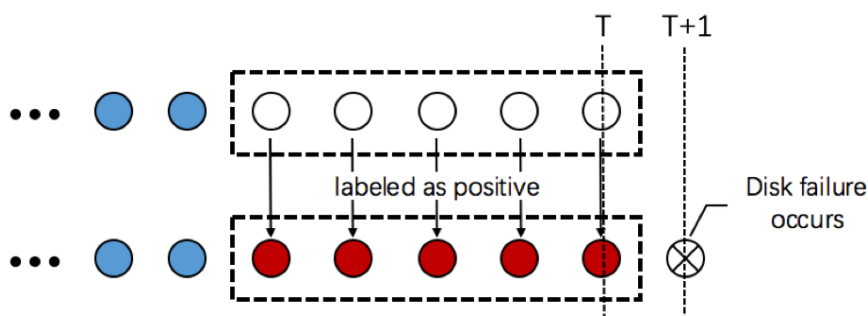


图 2-2 磁盘发生故障后样本的标记

假设数据中心中标志符为 i 的磁盘被标记为 D_i ，针对每一个磁盘 D_i 该方法为其创建一个定长的队列 Q_i ，用于暂存磁盘在最近一周时间内报告的 SMART 信息，这些存储与定出队列中的样本是保持未标记状态的，除非有新的样本到来或者所述磁盘发生故障。图 2-1 表示的是一个正常运行的磁盘，当有新样本采集来时样本的标记

情况，其中虚线框即为该磁盘的定长样本队列，虚线框中白色圆点表示的就是暂存的未标记样本，虚线框外蓝色样本表示标记为 **negative** 的样本。在 $T+1$ 时刻，伴随着新样本带来，由于样本队列已满，最左端的样本将出队列，而新样本则入队列，出队列的样本被标记为 **negative**，已标记的样本就作为训练样本用于在线模型的演化学习。

图 2-2 表示正在运行的磁盘发生故障之后样本的标记情况。如图所示， T 时刻仍在运行中的磁盘，在 $T+1$ 时刻发生故障，此时磁盘样本队列中的所有都将被标记为 **positive**，这些已标记的样本也会用于在线模型的训练。在未标记样本被标记之后并用于更新模型的同时，当前采集的新样本还会作为在线模型的预测输入。在线模型会根据新样本对磁盘的健康状态进行预测，若预测结果为 **positive**，则说明该磁盘有很大可能在接下来的一周内发生故障，技术人员可以根据该结果对该磁盘进行数据备份或者换盘；若预测结果为 **negative**，则说明磁盘暂时健康。具体的基于 ORF 的在线学习方法如算法 2 所示。

算法 2 基于 ORF 的在线学习方法

输入： 新采集的样本 \mathbf{x} ，如果 $y=1$ ， \mathbf{x} 为 NULL

磁盘标志符 i

磁盘当前的实际状态 y ，若磁盘正在运行则 $y=0$ ，若磁盘已故障则 $y=1$

输出： 对磁盘健康状态的预测结果 y'

```
1: // Model update phase
2: if  $y == 1$  then
3:   // Disk  $D_i$  failed
4:   while  $Q_i$  is not empty do
5:      $\mathbf{x}' \leftarrow \text{dequeue}(Q_i)$ 
6:      $\text{updateORF}(\langle \mathbf{x}', 1 \rangle)$  // call Algorithm 1
7:   end while
8:    $\text{deleteDisk}(D_i)$ 
9: else
10:  // Disk  $D_i$  is operating
11:  if  $\text{isFull}(Q_i)$  then
12:     $\mathbf{x}' \leftarrow \text{dequeue}(Q_i)$ 
13:     $\text{updateORF}(\langle \mathbf{x}', 0 \rangle)$  // call Algorithm 1
14:  end if
15:   $\text{enqueue}(Q_i, \mathbf{x})$ 
16: // Prediction phase
```

```
17:  $y' \leftarrow \text{predictORF}(ix)$ 
18: if  $y' == 1$  then
19:     //  $D_i$  is risky
20:     // Immediate data migration is recommended
21:     Trigger an alarm
22: end if
23: end if
```

2.3 本章小结

本章首先简单介绍了在线随机森林算法以及该算法的优势，接着提出了将在线算法运用于磁盘故障预测时会面临的两个挑战：在线模式下的样本标记问题和磁盘故障正负样本不平衡问题。针对这些挑战，本章具体讲解了可行的解决方案，并给出了优化后的 ORF 算法和基于 ORF 的在线学习算法。

3 在线磁盘故障预测系统设计

在线磁盘故障预测系统可以根据功能的不同分为数据预处理模块、模型训练模块、样本在线标记模块、模型在线更新模块和磁盘结果预测模块等五个模块。本章首先讲解系统的总体设计架构，以及总体设计中每个独立的模型的功能；接着具体针对五个模块的设计流程，模块的实现进行讲解。

3.1 系统总体架构设计

在线磁盘故障预测系统需要能够实际应用于大型数据中心，在提供高预测精度的同时还需要能保证系统的稳定运行。大型数据中心的应用环境非常复杂，且磁盘数量庞大，会达到百万级，预测系统会具有很高的吞吐量，考虑到磁盘故障预测的实施性要求，系统的设计需要充分考虑其运行效率。磁盘故障预测功能要求较高，且系统流程复杂，想要在短时间内完成数据的采集、数据清洗、模型训练、对所有磁盘健康状态的预测、在线对模型更新等过程是一项非常艰巨的任务。因此，本系统将各个功能模型分离，以最大限度的保障任务的并行执行，最大限度的提高系统的吞吐量。

本系统按照功能的不同以及处理的流程将整个系统分为五个子模块：数据预处理模块、模型训练模块、样本在线标记模块、模型在线更新模块和磁盘结果预测模块，如图 3-1 所示。数据预处理模块主要负责数据解析与规整，以及比较关键的数据标准化和特征选取过程，其中特征选取过程是结合历史数据以及预测模型在离线的环境下进行的，而数据的清洗和标准化则是在线进行的；模型训练模块使用历史数据训练预测模型，结合训练数据确定最优的模型超参数设定，比如 ORF 模型的参数：森林中树的棵树 T 、测试集合中测试的数量 N 、 λ_p 和 λ_n 、 $MinParentSize$ 和 $MinGain$ 等；样本在线标记模块负责陆续到来的样本的存储与标记问题，该模块将各磁盘最近一周内采集的 SMART 记录暂存起来，并根据记录采集的时间或者磁盘故障情况对样本进行标记和替换，具体的样本标记方案见 2.2.1 部分；模型在线更新模块使用已标记的样本对模型进行更新；磁盘结果预测模块对磁盘当前采集的未标记的数据进行预测，本课题的预测目标是磁盘是否会在未来一周内发生故障，因此如果预测结果为 1，则说明磁盘状态较差，需要进行及时的数据备份或者换盘。

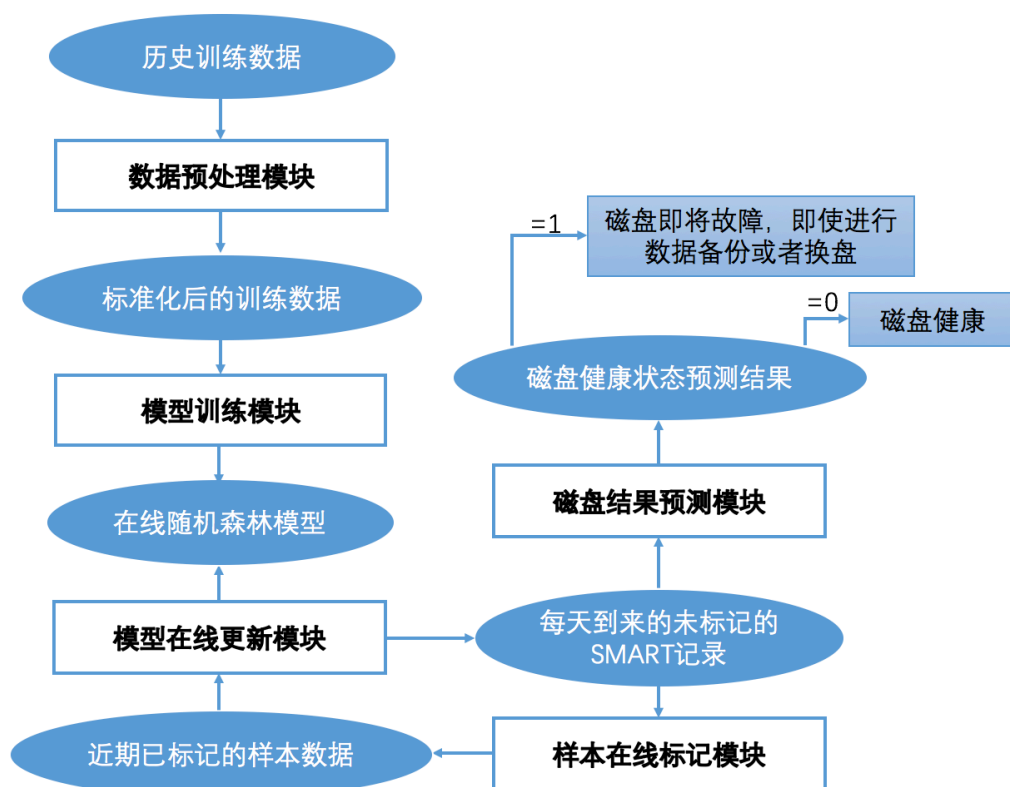


图 3-1 在线磁盘故障预测流程图

3.2 子系统功能模块设计

上一节将在线磁盘故障系统分为了五个功能模块，并分别简单介绍了每个模块的任务，本节将具体介绍每一个功能模型的设计方案。其中模型训练的具体实现参见 4.4 节，在线样本标记具体实现参见 2.2.2 节，这里不再赘述。本节主要对数据预处理模块进行讲解。



图 3-2 数据中心 SMART 数据采集架构

如图 3-2 所示，数据中心的每一台服务器上都需要部署采集 SMART 数据的 Agent，这些 Agent 将采集到的信息上报给后台，后台的数据预处理模块对这些数据进行解

析和规则，包括 SMART 属性的特征值的标准化。需要对特征进行标准化的原因是，每个特征的取值范围存在很大的差异，如果直接使用原始数值进行模型训练，则取值较大的特征会在模型训练中起到主导作用，即模型训练往往会偏向于这些取值较大的特征^[43]，而这种情况是不希望发生的。为了防止大取值特征主导模型的训练，所有特征均采用公式(3-1)进行标准化处理，将其原始数值映射到 0 至 1 的取值范围：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3-1)$$

其中， x 表示特征的原始数值， x_{\max} 和 x_{\min} 分别表示该特征取值的最大值和最小值，而 x' 则是标准化之后的值，直接用作模型训练的输入。需要注意，针对每一个特征，这一标注化公式需要同时作用于训练集和测试集，且针对同一个 disk model 的数据集，标准化公式中的 x_{\max} 和 x_{\min} 需始终保持一致，不能在标准化训练集的特征时使用训练集中该特征的最大值和最小值，而在标准化测试集时使用其他数值的最大值和最小值。

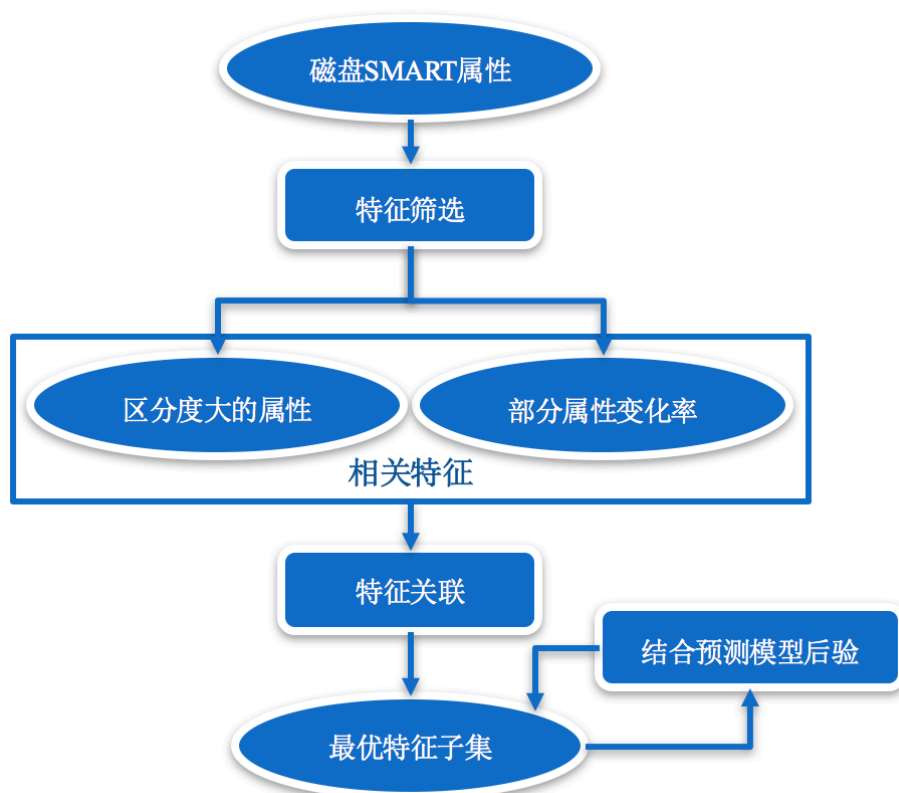


图 3-3 特征选取流程图

以上对数据进行标准化的过程是实时在线进行的，除此之外，还有一个至关重要的预处理过程需要进行——特征选取，这个过程通常是离线完成的，也就是说，只需要根据历史数据结合相关统计的方法以及预测模型，判断各个特征是否对故障预测起作用，然后选取出关键的特征集，仅使用关键特征集合中的特征进行模型的训练和后期磁盘健康状态的预测。之前的工作中都指出，样本特征集的好坏会直接影响磁盘故障预测模型的性能。有些特征随时间始终保持不变，另一些特征其正负样本区别很小，这些不相关特征会对磁盘故障的预测带来干扰，还会降低算法的计算效率。本系统数据预处理模块中的特征选取基于时间序列对比分析好盘/故障盘的各属性随时间的变化规律，并使用 rank-sum 测试，找出好盘与故障盘区分度大的 SMART 属性，作为相关特征。有些故障盘的部分特征在故障发生之前会出现明显下降趋势，因此本课题会选取出故障发生前有明显变化趋势的属性，将其变化率也作为相关特征进行考量。

文献[30, 31, 32]均使用 reverse arrangement test, rank-sum test 和 z-scores 等统计学方法进行特征选取，这些统计学方法只是选取出潜在有用的特征，未充分考虑到特征间的冗余性，且没有结合故障预测模型进行后验。本系统数据预处理模块中的特征选取在得到相关特征之后还会结合故障预测模型（本课题中使用的随机森林模型）的预测性能，考虑特征间的相关性，去除冗余特征，为预测模型提供更为精简有效的特征子集。特征选取过程如图 3-3 所示。

3.3 本章小结

本章主要讲解了在线磁盘故障预测系统设计方案，将系统分为数据预处理模块、模型训练模块、样本在线标记模块、模型在线更新模块和磁盘结果预测模块等五个模块分别进行讲解。

4 性能测试与评估

本章的目的在于评估基于 ORF 的在线学习方法的预测性能，并将 SVM、DT 和离线的 RF 算法作为对比方法进行比较分析。在磁盘故障预测领域，离线的 RF 模型已被证实要优于当前的许多的分类模型^[42]，因此如果本文提出的 ORF 模型能够在性能上逼近 RF 模型，那就能说明 ORF 模型的预测性能要优于当前的许多分类模型。此外，为了评估在线学习方法的自动更新的效果，本章还对模型在实际中的长期使用进行模拟，并将 ORF 模型的自动演化机制与 RF 模型的离线更新策略进行比较。在对模型进行训练和测试之前，本章首先介绍了所使用数据集的相关信息以及对数据进行预处理操作。

4.1 测试数据

为了能真实地评估本文提出的在线学习方法的预测性能，本课题使用的所有测试数据均来自于真实数据中心 BackBlaze。从 2013 年开始，BackBlaze 每天都会对其数据中心中的每一个正在运行的磁盘记录一条快照，这些快照包含磁盘报告的所有 SMART 属性值，以及快照记录的时间、磁盘的序列号和磁盘模型 (disk model) 等信息。BackBlaze 数据集一共包含 66,286,887 条样本记录，每条记录共 95 个字段，涵盖了 103,055 个磁盘的 SMART 数据，样本的记录时间窗口为"2013-04-10 "至"2017-03-31 "，跨度 47 个月，该数据集的样本字段名以及含义见表 4-1。

表 4-1 BackBlaze 数据集各字段含义

字段名	字段含义
date	样本记录时间，格式为 yyyy-mm-dd
serial_number	由磁盘生产商指定的磁盘序列编号
disk model	由磁盘生产商指定的磁盘模型编号
capacity_bytes	磁盘容量，以字节为单位
failure	如果磁盘还在运行，则该字段为 0；如果该记录为磁盘故障发生前记录的最后一条记录，则该字段为 1
SMART Stats	共 90 列数据，分表表示磁盘报告的 45 个不同的 SMART 属性的 Raw 值和 Normalized 值，可能存在 NULL 值

整个数据集中数十万块磁盘共涵盖了超过 30 个不同的 disk model。由于不同 disk model 的磁盘报告的 SMART 属性存在较大差异,即使都报告了同一个 SMART 属性,其属性的具体含义以及计量方式也可能不同,因此针对不同 disk model 的磁盘,需要单独训练磁盘故障预测模型。所以,本文不对 BackBlaze 的整个数据集进行训练建模,而是根据 disk model 字段将其划分为多个子数据集,在选择的子数据集上进行建模与评估。子数据集的划分结果见表 4-2,表中给出了部分子数据集包含的磁盘数量以及其中故障磁盘的数量。由于大部分 disk model 的子数据集只包含较少数量的故障磁盘,并且有些 disk model 的磁盘报告的 SMART 属性太少,因此综合多方面的考虑,本课题选取了两个拥有最多故障磁盘的 disk model 进行建模和测试,所选 disk model 和相对应的子数据集的统计信息见表 4-3。

表 4-2 BackBlaze 数据集依据 disk model 的划分结果

磁盘生产商	disk model	磁盘数量	故障磁盘数量
Seagate	'ST4000DM000'	36,541	1,996
Seagate	'ST3000DM001'	4,255	1,357
HGST	'HMS5C4040ALE640'	8,642	108
HGST	'HMS5C4040BLE640'	15,464	56
WDC	'WD60EFRX'	458	53

表 4-3 所选子数据集的统计信息

子数据集名称	STA	STB
disk model	'ST4000DM000'	'ST3000DM001'
磁盘容量 / TB	4	3
好盘数量	34,535	2,898
故障盘数量	1,996	1,357
样本采集时间跨度	39 个月	20 个月

本课题使用的测试数据集包含数千块故障磁盘,而工作[32]中使用的数据集仅包含数百块故障盘;此外,本课题的数据集的样本采集时间跨度达两年以上,而工作[32]中数据集的样本跨度仅两个月。这些优势使得在本测试数据上能更好的模拟真实数据中心中磁盘故障预测模型的长期使用,使得模拟的性能评估结果变得更加可靠,也更具说服力。

4.2 数据预处理

在使用测试数据进行分析之前，还需要先进行一些数据预处理的操作，以过滤掉冗余或者不相关的特征。这个数据预处理过程就是特征选取，该过程因为减少了训练模型时使用的特征数量，因而能大大减少模型训练和测试的时间，并且还在一定程度上能够提升模型的预测性能。

对于本课题选取的 STA 和 STB 这两个数据集，其包含的磁盘均报告了 24 个有效的 SMART 属性。每一个 SMART 属性都记录有两个数值：一个 6 字节的原始数值 (*Raw value*) 和一个 1 字节的标准化数值 (*Normalized value*)。由于标准化数值通常是由原始数值通过磁盘厂商指定的公司计算得出来的，有一些标准化数值会在这种公式转化后失去一定的精度，而其对应的原始数值则对于磁盘故障的检测更为敏感，因此每一个 SMART 属性的原始数值和标准化数值都应作为备选的输入特征，从而对于所选的两个 disk model，均有 48 个备选特征需要进行筛选。

表 4-4 通过特征选取的关键特征

ID#	SMART 属性名	Norm	Raw	rank
1	Read Error Rate	✓		13
5	Reallocated Sectors Count	✓	✓	3
7	Seek Error Rate	✓		7
9	Power-On Hours		✓	5
12	Power Cycle Count		✓	11
183	Runtime Bad Block		✓	8
184	End-to-End error	✓	✓	4
187	Reported Uncorrectable Errors	✓	✓	1
189	High Fly Writes	✓		10
193	Load Cycle Count	✓	✓	6
197	Current Pending Sector Count	✓	✓	2
198	(Offline) Uncorrectable Sector Count	✓	✓	9
199	UltraDMA CRC Error Count		✓	12

特征选取过程先对每一个特征运行威克森秩和测试 (Wilcoxon rank sum test)，以判断是否有足够的证据在 1%显著性水平下拒绝空假设，并得出该 SMART 特征在正负样本中的分布规律存在显著差异的结论。那些不论好盘还是坏盘，取值都保持

不变或者很接近的 SMART 特征在这一步骤中被过滤出来，这一步骤最终过滤掉 20 个对于磁盘故障预测毫无作用的特征。

接下来，特征选取过程对剩余的 28 个特征利用留一法，即每次选取其中的 27 个特征来训练随机森林模型，并比较这些模型的最终的故障检出率，从而判断每一个特征对于故障预测的贡献以及这 28 个特征中是否存在冗余。该步骤发现其中的 9 个特征是冗余的，集包含这些特征对最终预测模型的性能并没有任何提升，从而这些冗余特征也被过滤掉了。最终通过特征选取的特征一共有 19 个，其中包含 9 个标准化数值、10 个原始数值，具体特征见表 4-4。表中 rank 列表示 SMART 属性对于故障预测的贡献排名。一个 SMART 属性对于磁盘故障预测的贡献，可以通过计算使用全部 28 个特征与使用除去该 SMART 属性相关的两个特征后的 26 个特征训练的随机森林模型的预测性能差异来粗略判断。当除去某一个 SMART 属性的相关特征后，建立的预测模型性能减小越多，则说明该 SMART 属性对于故障预测的作用越大，其在表 4-4 中的 rank 排名越靠前。

4.3 性能评估指标

故障检出率 FDR 和误报率 FAR 被用于评估在线学习模型和其他用于比较的离线模型的预测性能。由于本研究中故障预测的目标是跟根据当前得到的 SMART 向量来预判磁盘是否会在接下来的一周之内发生故障，所以对于测试集中的故障磁盘，只有当该磁盘故障发生前一周之内采集的任意一个 SMART 样本被正确预测为 positive，那么这个故障磁盘才算被预测模型成功检出了。故障检出率 FDR 则定义为测试集中所有实际故障的磁盘中被正确预测为故障的磁盘所占的比例。对于测试集中的正常磁盘，只要该磁盘最近一周以外采集的所有的 SMART 样本中有任意一条被错误的预测为 positive，那么这个正常磁盘就算是误报的。误报率 FAR 定义为测试集中所有实际正常的磁盘中被误报为故障的磁盘所占的比例。

由表 4-5 可得到本实验中各指标的计算公式如公式(4-1)和公式(4-2)所示：

$$FDR = \frac{TP}{TP+FN} \quad (4-1)$$

$$FAR = \frac{FP}{FP+TN} \quad (4-2)$$

其中，TP 表示实际是故障磁盘测试结果也是故障磁盘；FN 表示实际是故障磁盘但测试结果是正常的，也就是漏报；FP 表示实际是好盘样本测试结果是坏盘样本，也就是误报；TN 表示实际是正常磁盘测试结果是正常的。注意，在 FDR 和 FAR 之间

通常存在着权衡，也就是说，随着 FDR 变大，FAR 往往也会随着增长，而预测性能的目标是达到尽可能接近 100% 的 FDR 的同时 FAR 不能太大。所以，在接下来的实验评估当中，比较不同预测模型的 FDR 时，会控制其 FAR 都保持一致。

表 4-5 四类预测结果

	测试结果为故障	测试结果为正常
实际故障(Positive)	TP(True Positive)	FN(False Negative)
实际正常(Negative)	FP(False Positive)	TN(True Negative)

4.4 ORF 模型性能评估

4.4.1 实验准备

为了评估基于 ORF 的在线学习模型，STA 和 STB 每个数据集中的所有磁盘被按照 7:3 的比例划分成训练集和测试集。其中训练集包含 70% 的正常磁盘和故障磁盘，而剩余 30% 的磁盘则包含于测试集。

为了更好的模拟模型在实际应用时的情形，对于训练集中的每个坏盘，仅有该盘故障发生前一周内采集的 SMART 数据被标记为 **positive**，因为本课题的预测目标是磁盘是否会在一周内发生故障，而该盘在故障发生前一周以外采集的 SMART 样本，在这些样本被采集时后一周之内磁盘并未出现故障，所以这部分样本均被标记为 **negative**。对于训练集中的每个正常磁盘，该盘最新一周时间内采集的样本是无法进行标记，即处于 **unlabeled** 状态，因为无法预判这些样本被采集后一周时间内是否该盘就发生故障了，而只有这最新的一周以外的样本才能标记为 **negative**，因为这些样本在采集后一周内磁盘均未发生故障。

值得注意的是，这种标记方法会将一些属于故障磁盘且具有明显故障特征但于是在故障发生前一周以外采集的样本标记为 **negative**。不过，这些类似于被误标记的 **negative** 样本在所有 **negative** 样本中只占很小一部分，因此这些误标记的样本被选中用于模型训练的几率很小。此外，本课题提出的 ORF 模型对于样本的误标记具有很好的鲁棒性，因此本标记方法中存在的误标记问题对模型预测性能的影响可以忽略不计，但是为了模拟模型在实际应用中在线使用时的情形，这种标记方法是必须的。

4.4.2 超参设定

由于故障磁盘只占有所有磁盘中很少的一部分，且每个故障磁盘报告的所有 SMART 记录中也只有很少一部分被标记为 positive，因此正负样本是严重不平衡的，这对预测模型的性能会带来很大的影响。如果使用原生的不平衡的数据集进行模型的训练，预测模型表现会很差，因为分类算法通常是以最大化总体的准确率为优化目标，也就是说，在不平衡的数据集上训练出来的模型会偏向于数量较多的那一类，从而大大限制了模型的预测性能。

为了解决这一问题，本课题针对 ORF 模型提出了两个超参数： λ_p 和 λ_n ，用于控制正负样本被模型选中而用于更新的几率；针对离线模型，本课题提出使用超参数 λ 来控制负样本的采样容量，以使得训练集中的负样本的数量跟正样本数量差异不会太大。这些超参数的选取会影响训练集中正负样本的比例，从而对预测模型的预测性能产生很大的影响，因此为了得到性能较好的预测模型，需要使用合适的超参数设置。本课题为了选取合适的超参数设置，基于使用不同的超参数设置得来的训练集训练预测模型，并在测试集上评估这些模型的预测性能。

在这部分的实验当中，随机森林中的树的棵树 T 被设定为 30，而决策树构建过程中结点分裂时使用的随机测试集中的测试数量 N 被设定为 5,000。随机森林中树的棵树 T 取更大的值时，预测性能没有明显的提升，反而会大大增加随机森林的训练与测试时间，因此 T 设定为 30 是合理的，不要设定为更大的数值。随机测试集中测试的数量设定为更大的值时，可能会稍微提升一些随机森林的预测精度，但是较大的测试数量设置会严重增加模型训练时间，所以是不可取的。对于 ORF 模型的其他一些参数的设置如下：MinParentSize $\alpha = 200$ ，MinGain $\beta = 0.1$ 。在接下来的讨论中，如没有特别声明，这些参数设置将保持不变。为了避免样本选取的随机性对模型预测性能的评估带来影响，该部分的全部实验都重复了五次，并且计算了预测模型 FDR 和 FAR 的均值与方差。

表 4-6 超参数 λ 对离线随机森林 RF 的影响

λ	STA		STB	
	FDR (%)	FAR (%)	FDR (%)	FAR (%)
1	98.22 \pm 0.25	11.88 \pm 2.62	92.26 \pm 0.27	6.46 \pm 0.54
2	99.02 \pm 0.31	2.33 \pm 0.95	88.77 \pm 0.38	1.35 \pm 0.18
3	98.16 \pm 0.74	0.76 \pm 0.17	85.45 \pm 0.54	0.73 \pm 0.06
4	94.58 \pm 0.64	0.05 \pm 0.04	82.61 \pm 0.65	0.67 \pm 0.04

5	92.00 ± 0.14	0.00	80.00 ± 0.55	0.66 ± 0.07
Max	35.14 ± 0.18	0.00	29.45 ± 0.76	0.00

正如在第 2 章中所讲解的，超参数 NegSampleRadio λ 被用于对含有较大数量的负类样本进行抽样，以得到与正类样本数量相当的负类样本，结合正类样本共同构成最终的训练集而用于模型训练。超参数 λ 对离线随机森林模型预测性能的影响见表 4-6。由该表可知，当不对负样本进行取样，而使用原始的训练集进行模型的训练（即表中 $\lambda = \text{Max}$ ）时，随机森林模型尽管能实现零误报，但其在 STA 数据集上的故障检出率只有 35%，而在 STB 数据集上故障检出率不到 30%，这样的预测性能远远没有达到实际使用中的需求。由于原始数据集中负样本的数量通常是正样本数量的成百上千倍，因此直接使用原始数据集训练的预测模型会严重的偏向负样本类，这就导致大量本具有故障特效的样本也被预测为 negative，从而使得模型的故障检出率收到了很大的限制。

一个可行的解决方案是，对占比太大的负样本类进行随机取样，而取样的容量就是通过超参数 λ 控制的。比如，当 λ 取值为 5 时，则最终的训练集中负样本的数量就是正样本数量的 5 倍。从表 4-6 中可以看出，随着 λ 的增大，实际用于训练离线模型的负样本数量逐渐减小，而预测模型也不再偏向于负样本类。这就使得模型的故障检出率 FDR 逐渐增加，而误报率 FAR 也会随着增加。较大的误报率会增加运维处理成本，因此误报率通常会有所限制。本课题对误报率的限制为 $\text{FAR} \leq 1\%$ ，因此对于离线随机森林模型，超参数 λ 被设定为 3；对于其他离线模型，如没有特殊申明，超参数 λ 也被设定为 3。

表 4-7 超参数 λ_n 对在线随机森林 ORF 的影响

λ_n	STA		STB	
	FDR (%)	FAR (%)	FDR (%)	FAR (%)
0.01	98.50±0.19	24.88±3.33	90.91±0.23	5.25±0.58
0.02	98.08±0.37	0.66±0.35	85.64±0.37	0.85±0.14
0.03	95.86±0.75	0.10±0.11	74.37±1.75	0.58±0.04
0.05	84.44±0.65	0.01±0.01	59.58±0.50	0.30±0.05
0.10	65.67±3.11	0.00	47.81±0.93	0.12±0.03
1.00	23.58±0.00	0.00	28.23±1.55	0.09±0.03

在线模型的实际应用中，训练样本是在预测模型投入使用的同时到来并作用于模型更新的，即陆续到来的、被标记后的 SMART 样本是在模型线上运行的同时作用于模型，从而进行模型的演化更新的。因此，离线模型所使用的样本取样对于在线情形下随时间陆续到来的样本不再适用。在线模式下，为了避免因为负样本远多于正样本而对预测模型的性能带来影响，本课题提出使用两个泊松分布函数分别作用于所有陆续到来的正、负样本，使得正、负样本具有不同的几率被选中而用于模型的更新。具体而言，正样本被作用于 $\text{Poisson}(\lambda_p)$ ，而 λ_p 可被设置为 1，这使得每一个后续到来的正样本具有约 70% 的几率被选中用于模型更新；而负样本则被作用于 $\text{Poisson}(\lambda_n)$ ，其中 λ_n 通常根据原始训练集中正负样本的比例进行设置，假设其被设置为 0.01，这使得每一个后续到来的负样本只有不到 1% 的几率被选中用于模型更新。因此，将 λ_n 设置的远小于 λ_p 可以让负样本被选中的几率远小于正样本被选中的几率，从而使得陆续到来的负样本中仅有一小部分被实际应用到了模型的更新上。

不同的 λ_n 选取会让负样本被选中的几率发生变化，进而决定在所有陆续到来的负样本中被实际选中而用于模型更新的样本数量，该参数与离线情形下对全体负样本进行随机取样时的取样容量作用相同。本课题中 λ_p 被设置为 1，而 λ_n 对在线随机森林的影响见表 4-7。由该表可以看出，随着 λ_n 的增加，模型的故障检出率逐渐减小，而误报率也是随之减小。这是因为， λ_n 的增加会使陆续到来的负样本被选中的几率变大，从而会有更多的负样本被选中用于模型的更新，使得模型更加偏向于负样本类。为了在保证误报率不超过 1% 的前提下实现尽可能高的故障检出率，本课题中 λ_n 被设置为 0.02。本课题所提出的基于 ORF 的在线学习方法能够实现 98% 的故障检出率，跟离线随机森林的预测性能相当，也远高于之前的研究工作提出的故障预测算法。

4.4.3 其他对比模型

除了离线随机森林模型外，本课题还将 ORF 模型与另外两个常用的分类模型进行比较：支持向量机模型和决策树模型。对于支持向量机模型，本课题使用被广泛使用的 LIBSVM 库，相关的详细介绍以及使用方法可参见文献[44]。该模型中的参数设置如下： $svm_type = C-SVC$, $kernel_type = RBF$ (radial basis function)。惩罚参数 $cost$ 和 RGB 核的参数 $gamma$ 可以用来调整模型以得到不同的 FDR 和 FAR 对。本课题针对这两个参数使用网格搜索 (grid search) 的方法选取能够使模型在 $FAR \leq 1\%$ 时实现最高 FDR 的参数设置。对于决策树模型，本课题使用 MATLAB 中提供的

fitctree 函数来创建分类树，相关介绍和函数使用方法详见页面³。该函数中的关键参数设置如下：*SplitCriterion* = gdi (Gini's diversity index), *MaxNumSplits* = 100。针对正负样本的不同的 *Weights* 可以用于该模型中 FDR 与 FAR 的调整。

4.4.4 ORF 模型评估

本课题根据已标记样本的时间戳来模拟训练数据逐渐到来的过程，而基于 ORF 的在线学习模型就是使用这些逐渐到来的训练样本进行演化学习。在模型学习演化的过程中，本课题每隔一个月就会对学习至当前这个月的 ORF 模型在测试集上的预测表现进行评估。与之进行比较的三个离线模型，则是使用当前这个月以及之前到来的全部训练样本进行离线地训练，然后在相同的测试集上进行测试评估。

图 4-1 和图 4-2 分别表示 ORF 模型以及三个离线模型在数据集 STA 和 STB 上的预测结果，图中所有的结点均满足 $FAR \leq 1\%$ 的限制。由图可以看出，在样本开始采集的前一两个月，所有模型的预测都较差。这是因为在样本开始采集初始时期，记录的故障磁盘太少，能够得到的有效 positive 样本数量受到了极大的限制，从而导致训练得到的模型因为缺乏足够的有效数据而发生欠拟合。也正是因为这个原因，许多模型在初始时期的误报率无法调整至 1% 以下，因此图中部分曲线并不是从第一个月开始的。

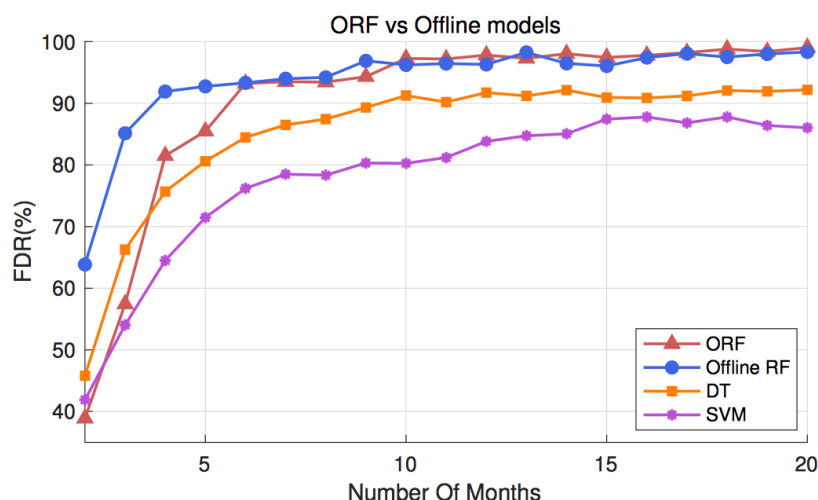


图 4-1 数据集 STA 上各模型的 FDR

³ Fit binary classification tree for multiclass classification. <https://cn.mathworks.com/help/stats/fitctree.html>

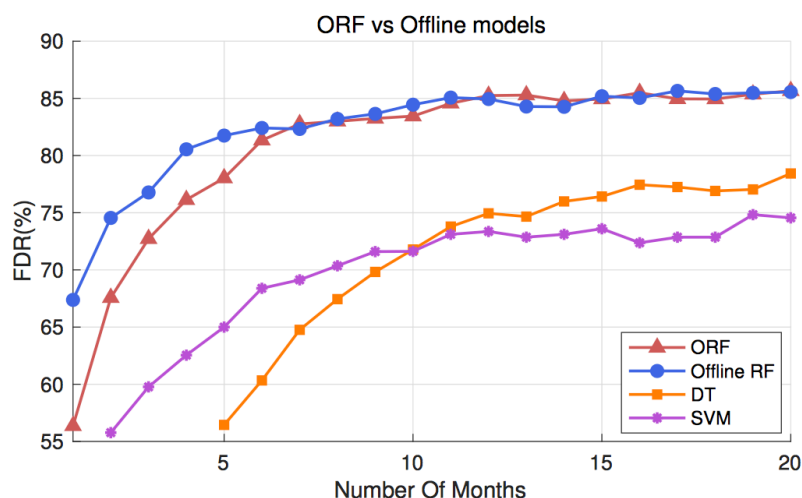


图 4-2 数据集 STB 上各模型的 FDR

随着样本采集时间的增加，故障磁盘也逐渐累积变多，能够得到的 positive 样本数量也随之增加。而伴随着有效数据的增长，训练得到的预测模型也越发准确，其故障检出率也得到快速提升。由图 4-1 和图 4-2 可知，三个离线模型中，随机森林模型的预测性能要远高于支持向量机和决策树模型，这也和故障预测方面的相关工作 [42] 的结论相同。本课题提出的 ORF 模型尽管在初始几个月的预测性能不及离线随机森林，但其可以在很短时间内迅速逼近了后者，然后保持与离线 RF 模型相差无几的预测性能。需要注意的是，ORF 模型在初始部署之后就无需每隔一段时间人为地进行模型的更新，因为其在在线上自动进行学习更新的；而离线模型则需要每隔一个月就使用先前采集的全部训练数据离线地从新训练一个新的模型。

从图 4-1 可以看出，在 STA 数据集上 ORF 模型能在第六个月之后实现稳定在 93-99% 之间的故障检出率。在第六个月的时候，训练数据集中故障磁盘只有 45 个，能够得到的有效 positive 样本也只有 299 个，这仅占全部正样本的 4%，尽管有效样本很有限，但是不论在线随机森林还是离线随机森林都能实现很好的预测性能，这说明本课题提出的基于 ORF 的在线学习方法不仅适用于大型数据中心，也能适用于有效数据很受限制的小型数据中心。由图 4-2 可知，在数据集 STB 上 ORF 模型的预测性能也是快速的在第七个月就逼近离线 RF 模型的预测性能，并实现稳定在 83-86% 之间的故障检出率。因此，实验表明本课题提出的 ORF 模型能够实现相较于之前的研究工作更好的预测性能。

4.5 模拟实际长期使用

ORF 模型不仅具有很好的预测性能,更重要的是其能够有效的避免离线模型的模型老化问题。正如之前所述,离线模型不具备自我更新的功能,模型训练出来之后除非以离线的重新训练为前提进行模型的更新,其在使用过程中是一成不变的。由于在磁盘故障预测中,所使用的属性集是磁盘 SMART 属性,而许多对故障预测起关键作用的 SMART 属性,其分布规律是随时间而发生变化的。这就导致预测模型尽管在初始时期具备很好的预测性能,但随着其使用时间的推移,模型预测性能会越来越差,也就是说,预测模型逐渐变得无法适应未来数据的模式。本课题提出在线学习方法与离线方法不同,在线模型能够在其投入使用的时候结合新到来的训练样本自动进行演化更新,这就使得在线模型能够始终适应当前的数据模式,而从根本上避免模型老化的问题。

本节通过模拟在线模型和离线模型的实际长期使用,比较在线模型的自动演化更新机制与离线模型的离线更新策略,来对在线模型自动更新机制的有效性进行评估。与 ORF 模型进行对比的是使用工作[32]中提出了两张离线更新策略进行更新的 RF 模型,这两种离线更新策略分别是 *accumulation* 更新策略和 *1-month replacing* 更新策略。前者定期(本课题每月更新一次)对模型进行更新,其使用从样本采集开始到当前一个月所采集的全部 SMART 样本来重新训练新的模型,而后者则只使用当前一个月之内采集的 SMART 样本进行模型训练,接着这些更新之后的新模型就被用于接下来一个月的样本预测。

SMART 样本的标记方法与 4.4 节中的标记方法相同。不同于前一节中将数据集 STA 或 STB 包含的样本按所属磁盘 7:3 的比例划分为训练集和测试集,本节将数据集 STA 或 STB 中的样本按样本的时间戳进行划分。例如对于数据集 STA,其样本采集时间跨度达 39 个月,该数据集中包含的全部样本均按照样本采集时间进行划分,每个月采集的样本划分在一起,这样就得到了 39 个子数据集。将 STA 中第 i 个月的子数据集标记 STA_i ,则 i 取值 1-39 之间整数。同样的,STB 数据集也根据样本时间戳被划分为 20 个子数据集,表示为 $\{STB_1, \dots, STB_{20}\}$ 。

对于不进行任何更新的使用离线模型,使用的训练集即为 STA 的前几个月的子数据集,而预测集则是之后的每一个子数据集。具体而言,对于数据集 STA,本课题使用前六个月采集的数据,即 $\{STA_1, \dots, STA_6\}$ 进行离线模型的训练,模型训练之

后,分别将该模型应用于 STA_7, \dots, STA_{39} 的预测,并评估其预测的结果。而对于数据集 STB,本课题使用前四个月的数据 $\{STB_1, \dots, STB_4\}$ 进行模型的训练。对于 *accumulation* 更新策略,在第 i 个月时,该策略使用前 i 个月采集的所有数据,即 $\{STA_1, \dots, STA_i\}$ 进行离线模型的训练,然后将该模型用于第 $i+1$ 个月的数据,即 STA_{i+1} 的预测;对于 *1-month replacing* 更新策略,第 i 个月时策略仅使用这一个月采集的数据,即 STA_i 进行离线模型的训练,然后用于 STA_{i+1} 的预测。对于本课题提出的在线学习方法,则不需要人为的进行模型更新,在线模型能够在运行过程中自动结合刚得到的训练数据进行演化更新。

本节中评估第 i 月时模型性能的方法跟前一节有所差异,这个差异主要体现在预测集上,前一节使用的预测集始终是固定的,而本节每月使用的预测集都有所差异,当然训练也不同。原因在于,前一节的目的是评估 ORF 模型相较于其他现存的预测模型运用到磁盘故障预测领域时的预测性能,即判断 ORF 模型是否适用于磁盘故障预测,而本节的目的则是评估模型在实际数据中心的长期使用中的表现,具体而言就是判断 ORF 模型是否可以很好的避免模型老化问题。

图 4-3 和图 4-4 显示了 ORF 模型和离线 RF 模型在实际的长期使用过程中的误报率,图 4-5 和图 4-6 显示了两个模型在长期使用中的故障检出率。由图中的紫色曲线可以看出,如果不对模型进行更新而长期的使用离线模型进行故障预测,模型的误报率会逐渐变大,于此同时模型故障检出率还会逐渐减小。这表明离线模型如果不加以更新,其在实际的长期使用中会逐渐失去其预测效果。值得注意的是,当模型误报率大于 5% 时,预测模型会带来大量的误报,这个数量甚至比真实故障的磁盘还要多,这会给数据中心的运维人员带来极大的负担。因此,如果使用离线模型进行磁盘故障的预测,定期更新模型是必不可少的。

由于刚训练出来的离线模型能够很好的适应当前数据的模式,而在较近的一段时间实现很好的预测性能,因此针对离线 RF 模型的 *accumulation* 更新策略可以很好的解决模型老化问题,因为其每个月更新一次模型并用于下一个月数据的预测。从图中看一看,在 STA 数据集上 *accumulation* 更新策略能将 RF 预测模型的误报率控制在 1% 附近,并且其每个月的故障检出率也能控制在 93-100% 之间;而在 STB 数据集上,该更新策略也能将 RF 模型的误报率控制在 2% 以下,且故障检出率在 80-90% 之间。图中可以看出,预测模型在每个月的故障检出率存在较大差异,而非稳定在一个数值附近,这是因为在每个月发生故障的磁盘数量差异很大,并且这些故障磁

盘当中不可预测的故障数量占比也不同。不可预测的磁盘故障在 SMART 数据上观察不到明显的故障特征，大都是因为机械或者电子组件问题而导致磁盘无法正常运行。

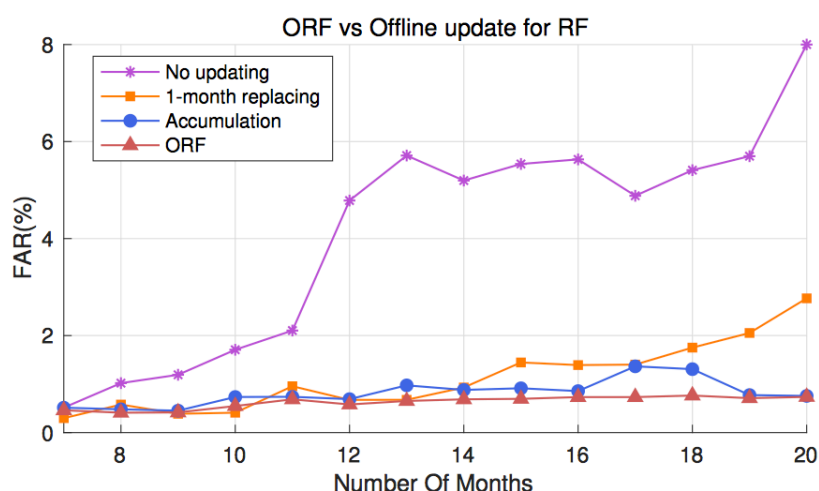


图 4-3 在 STA 数据集上各模型在实际长期使用时的 FAR

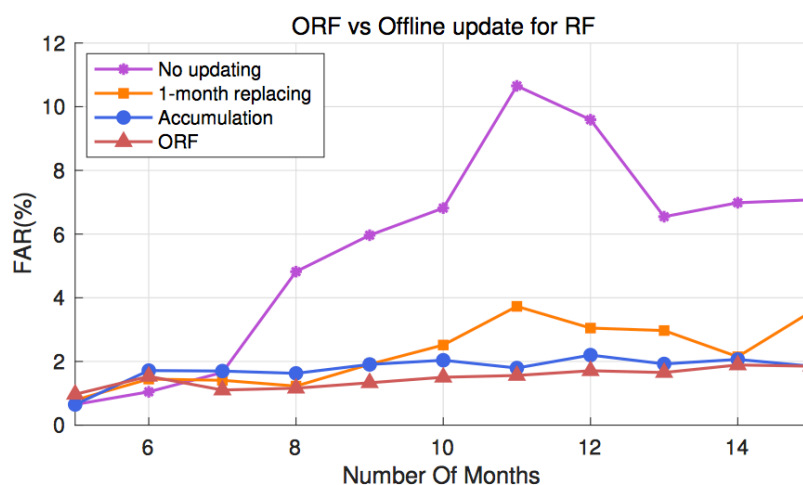


图 4-4 在 STB 数据集上各模型在实际长期使用时的 FAR

相较于 *accumulation* 更新策略，*1-month replacing* 更新策略表现要稍差一些，主要体现在稳定性上面。*replacing* 策略某一些月份会得到相较其他策略和模型最优的预测性能，但在大部分情况下，其预测性能不及 ORF 模型和 *accumulation* 更新的 RF 模型，并且 *replacing* 策略预测性能波动较大。一个可能的原因是该策略仅使用一个月采集的 SMART 数据进行模型的训练，而一个月故障的磁盘相对是非常有限的，从而训练集中有效 positive 样本数量也较有限，这会导致预测模型对磁盘故障信息的评估不具备好的鲁棒性。

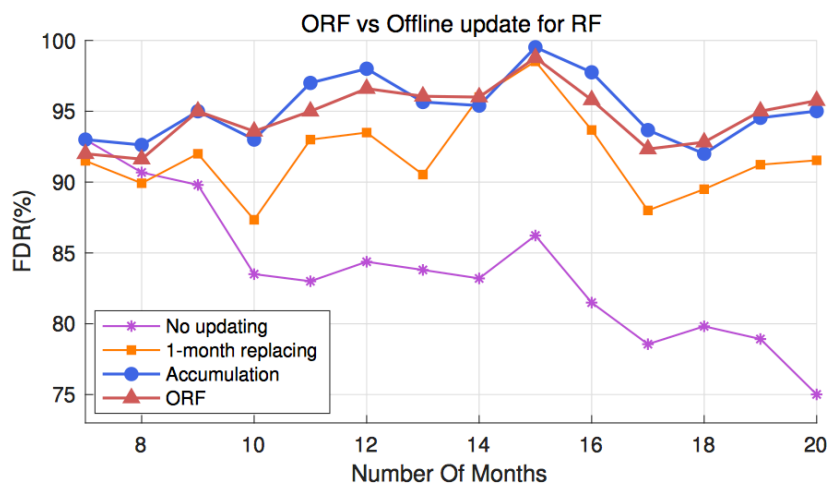


图 4-5 在 STA 数据集上各模型在实际长期使用时的 FDR

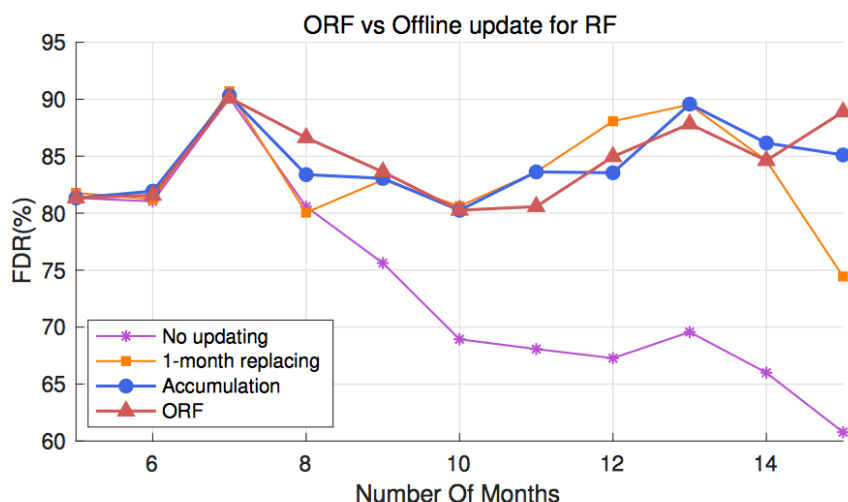


图 4-6 在 STB 数据集上各模型在实际长期使用时的 FDR

与使用这两个更新策略定期进行更新的 RF 模型相比，本课题提出的 ORF 模型无需每个月重新训练一个新的模型，但却能在维持更低的误报率时，实现跟 RF 模型相当的故障检出率。这进一步证实了 ORF 模型的可用性。

4.6 本章小结

本章首先对测试数据的来源和相关的统计信息进行了说明，接着讲解了对测试数据进行的预处理操作，即特征选取，给出了真正适用于磁盘故障预测的关键特征集。本章的重点在于对基于 ORF 的在线学习方法进行性能评估，与之进行比较的离

线 RF、SVM 和 DT 等常用分类学习方法。实验结果表明，ORF 模型能够实现与 RF 模型相差无几的预测性能，且远远优于 SVM 和 DT 模型，这证实了 ORF 模型在磁盘故障预测领域的适用性。此外，为了评估在线学习方法的自动演化更新机制的效果，本章还模拟了该模型在实际数据中心中的长期使用情况，实验表明在线学习方法能后很好的避免离线模型会遇到的模型老化问题，并且实现与离线 RF 模型相当的故障检出率。

5 总结及展望

数据中心存储系统规模的不断扩大使得原本偶发的磁盘故障变得越来越频繁，磁盘故障不仅会造成集群服务宕机更会导致敏感数据的丢失，这无疑会对数据中心的正常运营带来很大的影响。为了提升存储系统可靠性，传统的方法是在磁盘故障发生之后立即通过副本和编码等手段进行处理，比如常用的 RAID 技术、Google 的随机副本技术等。然而这些被动容错技术，大都会对磁盘的读写性能带来影响，甚至还会降低存储系统的整体性能。

主动容错技术则是在磁盘故障发生之前结合磁盘历史数据对其健康状态进行预测，使得用户得以根据预测结构提前采取诸如换盘或者数据迁移等措施来避免故障的发生，从而在不影响存储系统整体性能的前提下极大的提升数据可靠性。主动容错技术虽然具有很好的应用前景，但其在实际使用中的预测性能却不够理想。为了提升故障预测的精度，大量基于统计学或机器学习的方法的磁盘故障预测技术被提出，这些技术尽管实现了显著的性能提升，但他们大都没有考虑到预测模型在实际的长期使用中会遇到的模型老化问题。现有的研究成果都是在离线模式下训练预测模型，这些模型通常被假设为训练之后就不再改变。然而磁盘 SMART 数据的分布规律是随时间变化的，这就导致当前训练的预测模型尽管起初具有很好的性能，但随着时间的推移，其预测性能越来越差，即预测模型无法适应未来的数据模式。

为了得到适用于实际长期使用的预测模型，本文提出使用在线学习的方法进行预测建模。将在线学习运用到磁盘故障预测会面临两个挑战：1) 如何在线地对陆续收集到的 SMART 样本进行标记？2) 如何克服正负样本不平衡对预测模型性能的影响？对于前者，本文提出了一个自动在线标记方法，该方法将每一块磁盘最近 7 天（也可以是其他时间窗口）内采集的样本暂存起来，这些暂存着的样本将保持未标记状态，直到磁盘故障真实发生了，或者有新的样本到来。当一块磁盘故障之后，对应于该盘的最近 7 日的样本将被标记为 **positive**；而伴随新样本的到来，最先采集的样本将被这些新样本所替换，而这个替换出来的样本则被标记为 **negative**。一旦样本被标记之后，就可以作为训练样本输入到在线学习模型进行模型更新。对有第二个挑战，本文对 **online bagging** 进行改进，提出使用两个不同参数的柏松分布函数分别作用在陆续到来的正负样本之上，使得负样本相对正样本具有更小的概率被选中

而实际地用于模型更新。这样，真实作用于在线模型的负样本数量就可以被控制到与正样本数量相当的程度，从而解决样本不平衡的问题。

本文基于在线随机森林算法具体的实现了一个用于磁盘故障预测的在线学习模型。相较于之前的研究工作中提出的离线模型，基于 ORF 的在线模型具有内存需求更低、对于标记噪音具有更好的鲁棒性、模型可解释性更好以及更好的预测性能等优势。在来自于真实数据中心的数据集上的评估实验表明，本文提出的基于 ORF 的在线学习模型能够实现稳定在 93-99% 之间的故障检出率，并且误报率更低。更重要的是，ORF 模型能够实时地结合陆续采集的训练数据进行自动地更新演化，因此该模型对于 SMART 数据分布规律的动态变化具有很好的自适应性，这使得本文提出的在线学习方法更够彻底摆脱模型老化的问题。为了评估模型自动演化的效果，本文还模拟了其在实际数据中心中的长期使用。实验结果表明，在长达两年的实际使用中，相较于定期更新的 RF 模型，ORF 模型能够在维持更低误报率的同时，实现差不多的故障检出率。因此，本文提出的基于在线学习的磁盘故障预测技术适用于实际的长期使用，且预测性能很好。

未来的研究工作，将在现有工作基础上，在以下三个方面进行完善：1) 用于建模和评估的数据方面，现有工作使用的两个数据集来自于相同的磁盘生产商，而不同磁盘生产商的磁盘支持的 SMART 属性存在较大差异，且相同 SMART 属性的含义也存在较大差异。因此，在未来的工作中，会使用更多的、来自于不同生产商的磁盘数据进行模型的分析与评估；2) 使用的在线算法方面，本文仅基于在线随机森林建立了在线模型，与之进行比较的 RF、SVM 和 DT 模型均是离线模型。未来可以尝试使用更多的在线学习算法，进一步提升预测模型的预测性能和稳定性；3) 模型训练时的参数调优方面，由于时间限制，本文在实验评估阶段对各模型的超参调整时选取的粒度较大，尽管得到很很不错的预测性能，但未来仍可以使用更小的调参力度，以得到预测性能更好的模型超参选择。

最后，虽然本文是面向磁盘故障预测场景提出了一个在线学习方案，但是值得一提的是，该方案还能够很好的应用于更多的检测场景，只要这些场景下，训练训练数据是逐渐到来的，或者预测模型对于未来数据模式的适应性至关重要时，本文提出的在线学习方案均能起到一定的作用。

致 谢

当梧桐树又开始在校园里郁郁葱葱的时候，我完稿了此篇毕业论文。三年的研究生生涯，也由此进入尾声。回首这漫长而又短暂的三年时光，我心中充满了感恩与不舍。

尤记起第一次踏进东五楼实验室，严谨的科研氛围、舒适的学习环境、齐全的实验设备，让我对接下来的科研生涯充满了向往与憧憬。感谢金海老师，带领 CGCL 这个大家庭团结一致，不断发展壮大。在这里，我聆听了老师的谆谆教诲，获得了学长的热心帮助，领悟了 CGCL 实验室的科研精神，还收获了珍贵的友情。也是在这里，我第一次知道了云计算，接触了数据中心智能运维；第一次学习和掌握了各种机器学习和深度学习的算法，搭建了一个又一个学习环境，对各种数据进行了一次又一次的挖掘学习；还第一次在自己的努力和吴松老师的帮助下完成了一篇国家专利的申请。

科研的道路充满挑战，一路上充实与迷茫交织，灵感与瓶颈交织，快乐与孤独交织。感谢我的导师胡侃老师对我论文的关心和在学术探究上的指导，时常督促我的科研进度，胡老师对我的生活也有着体贴入微的关心，令我感到了师门的温暖。感谢我的指导老师吴松老师对我学业和生活上莫大的帮助。从课题的开题到毕业论文答辩，吴老师一直给予我许多指导，在我遇到瓶颈的时候耐心地帮我理清思路，明确研究方向。另外，吴老师还带领我参与了 863 计划云服务与管理平台核心软件及系统的研发工作，让我在实践中得到了充分锻炼，不仅让我了解了科研项目的整体流程，磨砺了我做学术报告的语言组织能力，见识了许多高校和科研机构的大牛，更重要的，让我对学术研究产生了浓厚的兴趣。此外，还要特别感谢后期才加入本课题的肖江老师。肖老师非常认真仔细地帮我修改论文，逐字逐句的帮我指出文中大大小小的问题，无比耐心地帮我整理论文的逻辑结构，给予我很多新鲜的思路，在工作日之外依然花费自己的私人时间给予我指导和帮助。最后，还要感谢我的师兄易宇声博士，易博在自身科研压力较大的情况下，还时时刻刻都在关心着我的科研进度，经常督促我的实验进度，与我共同探讨课题终于到难题，热心地帮我解答困惑，还在论文写作、实验方案设计上给了我很多前瞻性的意见，让我少走了许多弯路。

除此之外，在生活上，我还要感谢我的父母，感谢他们多年的辛劳与无私的付出，让我得以安心在学校里学习、成长。感谢 CGCL 大家庭每年举办的年会和元宵晚会，丰富了我的业余生活，给了我第二个“家”，让我得以在严谨而又温馨的环境里不断发展自我。感谢平日里一起玩耍的小伙伴陶晟、李浪、陈洋等等，让温暖的友情陪伴我度过了三年的校园时光，给我留下一份学生时代的美好回忆。最后还要感谢我的精神支柱顾萌萌同学，是她在我最困难的时候，给予了我继续奋斗的勇气和力量，是她一次又一次地将我从泥潭中解救出来，感谢萌萌同学这三年的付出和善良。

临近毕业，我将带着 CGCL 精神，迈入新的征程。在此，愿老师和同学们在以后的日子里万事顺意，愿自己不忘初心，学以致用，万里归来年愈少。

参考文献

- [1] B. Schroeder and G. A. Gibson. Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?[C]. In: Proceedings of 5th USENIX Conference on File and Storage Technologies, FAST'07, 2007: 1-16
- [2] Ioannis nousakis, Sriram Sankar, Gregg McKnight, Thu D. Nguyen, Ricardo Bianchini, Environmental Conditions and Disk Reliability in Free-Cooled Datacenters[C]. In: Proceedings of 14th USENIX Conference on File and Storage Technologies, FAST'16, 2016: 53-65
- [3] K. V. Vishwanath and N. Nagappan, Characterizing cloud computing hardware reliability[C]. In: Proceedings of the 1st ACM symposium on Cloud computing, ACM, 2010: 193-204
- [4] Haibo Mi, Huaimin Wang, Gang Yin, et al. Online Self-reconfiguration with Performance Guarantee for Energy-efficient Large-scale Cloud Computing Data Centers[C]. In: Proceedings of the 7th IEEE International Conference on Services Computing. Miami, USA: IEEE Computer Society, 2010. 20-28
- [5] 刘军平. 磁盘存储系统可靠性技术研究[D]. 武汉: 华中科技大学图书馆, 2011.
- [6] L. N. Bairavasundaram, G. R. Goodson, S. Pasupathy, et al. An analysis of latent sector errors in disk drives[J]. ACM SIGMETRICS Performance Evaluation Review, 2007, 35(1): 289-300
- [7] L. N. Bairavasundaram, G. R. Goodson, B. Schroeder, et al. An analysis of data corruption in the storage stack[J]. ACM Transactions on Storage, 2008, 4(3): 1-28
- [8] E. Brewer, L. Ying, L. Greenfield, et al. Disks for Data Centers. Keynote talk for FAST'16, Technical report, Google, Feb 2016.
- [9] P. Eduardo, W. Wolf-Dietrich, B. Luiz. Failure trends in a large disk drive population[C]. In: Proceedings of the 5th USENIX Conference on File and Storage Technologies. San Jose, USA: USENIX Association, 2007: 17-23
- [10] Yuanhui Sun, Zongshui Xiao, Dongmei Bao, et al. An architecture model of management and monitoring on cloud services resources[C]. In: Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering. Chengdu, China: IEEE Computer Society, 2010: 30-37
- [11] P. M. Chen, E. K. Lee, G. A. Gibson, et al. RAID: High-performance, reliable secondary storage[J]. ACM Computing Surveys, CSUR, 1994, 26(2): 145-185

- [12] H. Gobiof, S. T. S. Ghemawat. The Google file system[C]. In: Proceedings of the 19th ACM Symposium on Operating Systems Principles, SOSP'03, 2003: 29-43
- [13] R. Van Renese, F. B. Schneider. Chain replication for supporting high throughput and availability[C]. In: Proceedings of OSDI'04, 2004: 91-104
- [14] 曹小坤. 边缘云计算体系结构及数据迁移方法研究[D]. 西安电子科技大学, 2013.
- [15] Bruce Allen. Monitoring hard disks with smart. Linux Journal, 2004.
- [16] Joseph F. Murray, Gordon F. Hughes, Kenneth Kreutz-Delgado. Machine learning methods for predicting failures in hard drives:A multiple-instance application[J]. Journal of Machine Learning Research, 2005, 6(1): 783-816
- [17] T. Mitchell 著, 曾华君等译. 机器学习. 北京: 机械工业出版社, 2008: 23-56
- [18] 董勇, 蒋艳凰, 卢宇彤等人. 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
- [19] G. Hamerly and C. Elkan. Bayesian approaches to failure prediction for disk drives. In: Proceedings of the 18th International Conference on Machine Learning, ICML'01, 2001: 202-209
- [20] G. F. Hughes, J. F. Murray, K. Kreutz-Delgado, et al. Improved disk-drive failure warnings. In: IEEE Transactions on Reliability, 2002, 51(3): 350-357
- [21] J. F. Murray, G. F. Hughes, and K. Kreutz-Delgado. Hard drive failure prediction using non-parametric statistical methods[C]. In: Proceedings of the ICANN/ICONIP, 2003.
- [22] J. F. Murray, G. F. Hughes, and K. Kreutz-Delgado. Machine learning methods for predicting failures in hard drives: A multiple instance application[J]. Journal of Machine Learning Research, 2005, 6(1): 783-816
- [23] Y. Tan, X. Gu. On predictability of system anomalies in real world[C]. In: Proceedings of the 18th Annual IEEE/ACM International Symposium on Modeling, Analysis Simulation of Computer and Telecommunication System, 2010.
- [24] Y. Zhao, X. Liu, Gan S, et al. Predicting disk failure with HMM-and HSMM-based approaches[C]. In: Proceedings of the 10th International Conference on Data Mining, 2010: 390-404

- [25] Y. Wang, Q. Miao, and M. Pecht. Health monitoring of hard disk drive based on mahalanobis distance[C]. Prognostics and System Health Management Conference (PHM-Shenzhen), Shenzhen, China, 2011: 1-8
- [26] Y. Wang, Q. Miao, E. W. Ma, et al. Online anomaly detection for hard disk drives based on mahalanobis distance[J]. IEEE Transactions on Reliability, vol. 62, no. 1, 2013: 136-145
- [27] Y. Wang, E. W. Ma, Tommy W. S. Chow, 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
- [28] 刘景宁, 饶国林, 冯丹. 一种基于S.M.A.R.T的保障RAID数据高可靠性的方法[J]. 计算机工程与科学, 2007(05): 109-111
- [29] 张超. 高性能磁盘阵列自修复技术研究[D]. 长沙: 国防科学技术大学, 2008.
- [30] 胡维. 基于智能预警和自修复的高可靠磁盘阵列关键技术研究[D]. 长沙: 国防科学技术大学, 2010.
- [31] B. Zhu, G. Wang, X. Liu, et al. Proactive drive failure prediction for large scale storage systems[C]. In: Proceedings of IEEE Conference on Masive Data Storage Systems and Technologies, 2013: 1-5
- [32] J. Li, X. Ji, Y. Jia, B. Zhu. Hard Drive Failure Prediction Using Classification and Regression Trees[C]. In: Proceedings of the 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN'14, 2014.
- [33] C. Xu, G. Wang, X. Liu, et al. Health Status Assessment and Failure Prediction for Hard Drives with Recurrent Neural Networks[J]. In: IEEE Transactions on Computers, vol. 65, no. 11, 2016: 3502-3508
- [34] J. Li, R. J. Stones, G. Wang, X. Liu, Z. Li, M. Xu. Hard drive failure prediction using Decision Trees[C]. In: Reliability Engineering and System Safety, 2017: 55-65
- [35] J. Li, R. J. Stones, G. Wang, et al. Being accurate is not enough: New metrics for disk failure prediction[C]. In: Proceedings Symposium on Reliable Distributed Systems (SRDS), 2016: 71-80
- [36] V. Agarwal, C. Bhattacharyya, T. Niranjana, et al. Discovering rules from disk events for predicting hard drive failures[C]. International Conference on Machine Learning and Applications, 2009: 782-786

- [37] O. Fontenla-Romero, B. Guijarro-Berdin ĹČas, D. Martinez-Rego, et al. Online machine learning[C]. In: Efficiency and Scalability Methods for Computational Intellect, 2013: 27-54
- [38] N. C. Oza and S. J. Russell. Online bagging and boosting[C]. In: Proceedings Artificial Intelligence and Statistics, 2001: 105-112
- [39] A. Saffari, C. Leistner, J. Santner, et al. On-line Random Forests[C]. In: 12th International Conference on Computer Vision Workshops (ICCV), 2009: 1393-1400
- [40] L. Breiman. Out-of-bag estimates. Technical report, 1996.
- [41] L. Breiman. Random forests. Machine Learning, October 2001. 45(1):5-32
- [42] F. Mahdisoltani, I. Stefanovici, and B. Schroeder. Proactive error prediction to improve storage system reliability. In: Proceedings of the 2017 USENIX Annual Technical Conference (USENIX ATC'17), 2017: 391–402
- [43] J. Dinan, L. Brain, S. Ponnuswamy, et al. Scalable work stealing. in: Proceedings of the 2009 International Conference on High Performance Computing Networking, Storage and Analysis, 2009. 53-61
- [44] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. In: ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, 2011: 1-27

附录 1 攻读硕士学位期间发表的论文

[1] Jiang Xiao, Zhuang Xiong, Song Wu, et al. Disk Failure Prediction in Data Centers via Online Learning [C]. In ICPP 2018: 47th International Conference on Parallel Processing, August 13–16, 2018, Eugene, OR, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3225058.3225106>

附录 2 攻读硕士学位期间发明的专利

[1] 吴松, 金海, 熊壮, 胡侃. 一种面向磁盘故障检测的 SMART 阈值优化方法. 申请号: 201810006503.8, 专利申请日期: 2018.03.21, 单位: 华中科技大学

附录 3 攻读硕士学位期间参与的科研项目

- [1] 腾讯-华中科技大学合作项目，数据中心磁盘故障分析与预测，2015.9-2016.9
- [2] 国家高技术研究发展计划（863 计划）信息技术领域重点项目，云服务和管理平台共性基础核心软件，2014.1-2016.12