

**云计算与虚拟化论文读后感**

|  |  |
| --- | --- |
| 姓 名： | 梁一飞 |
| 学 院： | 计算机学院 |
| 专 业： | 计算机专业 |
| 班 级： | CS1706 |
| 学 号： | U201714762 |
| 指导教师： | 吴松 |
| 所选题目： | Disk Failure Prediction in Data Centers via Online Learning |
|  |  |

|  |  |
| --- | --- |
| 分数 |  |
| 教师签名 |  |

2020 年 6 月 10 日

# 一.阅读体会

在本文中，作者提出了一种使用在线学习方法进行主动容错的新方法。在线模式下训练磁盘故障预测模型存在两个主要挑战：1）如何实时标记顺序采集的样本 2）如何克服健康磁盘和故障磁盘的高度不平衡分布。对于前者，作者引入了一种自动在线标记方法，该方法可以暂时保留最近几天收集的样本以免标记，直到发生磁盘故障或新样本到达为止。对于后者，作者提出了两个泊松分布来模拟正样本和负样本的顺序到达，并且负样本具有较小的机会被模型选择进行更新。因此，用于训练的阴性样本可以与阳性样本的大小相比。此外，作者还建立了基于ORF的预测模型来支持他使用的方法。与以前的工作提出的离线模型相比，作者的ORF模型具有较低的内存需求，良好的标签噪声鲁棒性和更好的预测性能。对真实数据集的实验表明，作者的ORF模型可以在合理的低FAR的情况下实现93-99％的FDR。此外，作者的ORF模型在实际使用中具有特殊的优势，因为它可以随着数据的顺序到达实时执行自动更新，因此高度适应训练数据的动态分布。总而言之，作者的在线学习方法可以摆脱模型老化问题，并且不需要脱机更新。

为了评估方法的有效性，作者模拟了ORF模型的实际长期使用，并将其与离线RF模型的更新策略进行了比较。结果表明，作者基于ORF的算法可以保持合理的较低FAR，同时与这些定期更新的RF模型实现可比的FDR。因此，作者证明了在线学习方法在实际长期使用中保持稳定的预测性能的能力。最后，尽管作者的方法是在Seagate的两个磁盘模型上构建和评估的，但是只要支持SMART，它就可以轻松地应用于其他磁盘模型和制造商。此外，作者的方法应该适用于各种检测应用，在这些应用中，训练数据可以顺序可用，或者算法必须自动适应新的数据模式。

# 二.课程体会

整个课程上下来轻松愉快，正如老师所言，这个课纯兴趣课，所以没有太大压力可言,同时老师上课教学资料丰富，听的话可以收获很多，拓宽了知识面。而且结课形式也比较新颖,无论是像老师说的想"水"也好,想学也好,最后都各有所得,虽然这门课没有像其他课程那样布置困难的学习任务,但是不影响这是一门好课程,对于我来说,有的课程设置了太困难的任务而缺乏详尽的指导,最后真有所得吗?所得甚少。而云计算这门课,就是不强迫你学,但是做到了"开卷有益",是一门好课。