

主席先生，尊敬的客人，早上好！我叫刘恒宇。我来自西北工业大学。我很高兴能参加这次会议。今天，我想发表我的论文“基于不平衡数据的系统故障诊断方法研究”。

我的演讲将涉及四个方面。绪论、提出的框架、案例研究，最后一部分是总结和未来研究。

1、首先介绍了本文的研究背景。2.本页说明了分类技术在诊断中的重要性。3、第一部分是信息采集与处理系统。一般来说，不同的属性数据是由多个传感器采集的，需要进行预处理。然后它将被发送到分类系统。利用分类模型得到能够实现故障诊断的故障类别。4、具有异常信息、故障信息和预测信息。

1、在故障诊断方面，本文采用一种新的分类模型来获取故障类别，为维修策略提供有用的信息。2、利用历史数据训练分类模型。然后，当新的监测数据到达时，经过训练的分类模型将给出故障类别，完成故障诊断。3、故障类别的识别能够准确地实现复杂系统的诊断和适当的维护，对实际应用中的安全可靠运行具有重要意义。

本研究中提出的框架包括四个部分。

支持向量域描述模型是一种单类分类算法。其基本思想是通过非线性映射将原始样本空间转化为新的高维特征空间，并在高维空间中找到最优超球面。因此，它可以尽可能包含所有样本，并称量最大样本数以及球体的最小半径。

在我们的方法中，我们使用sigmoid函数来修改svdd的输出，如公式所示，这种分类是一个递归过程：首先，我们初始化svdd算法的参数；其次，我们需要训练超球面模型并结合SigMID函数得到SVDD算法的概率输出值；3：然后计算模型上的AUC值，4：最后，我们使用差分进化算法迭代优化算法参数，并使模型的AUC值最大化。数据集。

同时，随机森林模型是一种单类分类算法。随机林是套袋法的推广。该方法的主要思想是：在训练数据集上采用bootstrap抽样，训练子集由随机抽样和带回退次数组成，然后在训练子集上训练子分类器模型，重复抽样和训练次数得到子分类器模型。

优化的随机森林模型优化的RF算法的关键步骤如下：1：首先，我们初始化RF算法参数，例如树的数量，树的深度和最大树特征的数量等。2：第二，结合分层抽样方法，随机生成100个CART子树。3：此外，我们计算CART子树的AUC值，并使用30%的分位数确定AUC值的阈值参数；4：然后，我们使用分层采样方法生成训练CART子树，计算CART子树的AUC值，如果AUC值小于阈值，则删除CART子树。5：然后，我们重复CART子树的生成，直到满足终止条件为止。6：我们还选择投票方式来融合多个子树模型；7：最后，我们使用迭代优化方法，差分进化算法找到最优算法参数，使得该模型在数据集上的AUC值最大。

梯度提升决策树是Boosting方法的扩展。它使用决策树作为模型训练的子分类器，属于Boosting Tree模型。

在本文中，邻域清洁规则（NCL）用于对大多数样本进行欠采样。NCL算法结合了样本的分布和结构特征，有效保留了大多数样本的信息，比传统的欠样本方法更准确。优化后的GBDT算法的关键步骤如下：1：首先，我们多次运行NCL算法，直到数据集的不平衡度达到要求为止。2：然后，通过迭代优化方法（如差分进化算法）找到梯度提升树的最优模型。

自适应不平衡分类1：该算法首先结合one-vs-rest (ovr) 方法将不平衡分类问题分解为二值分类问题。状态的二进制分类数据集表示为。2:接下来在数据集上用优化的SVDD算法训练子模型。3:然后在数据集上用优化的射频算法训练子模型。4:然后在数据集上结合NCL算法和GBDT算法训练子模型。5:接下来，我们使用加权平均法来融合子模型，然后。然后输出融合模型。6:然后，我们使用交叉验证集来确定状态的最终模型。7:最后遍历设备状态的数据集，重复步骤2~6，得到最优模型。

自适应不平衡分类方法趋于获得更准确的分类结果。在介绍了所提出的方法之后，现在通过案例研究来验证所提出方法的有效性。下一部分是案例研究。我们选择2015 PHM数据挑战数据集作为案例研究。该数据集是典型的不平衡分类问题。该数据集基于美国PHM协会在2015年发布的设备运行数据集。该数据集包含2009年至2012年间60台设备的运行数据。数据运行可分为两类。第一类是操作信息，它是

设备的操作状态信息。第二类是故障信息，记录设备运行过程中的故障信息，包括开始时间和结束时间以及故障类型。该数据集旨在提前检测工厂故障事件。给定过去标记的故障类型，我们的目标是根据过去的预测类型的未来故障事件及其发生的时间。

在本文中，选择算法作为性能评估指标，定义为GCS。符号 TP 表示正确识别了非故障开始时间。符号 TN 表示已正确识别故障开始时间；符号 FN 表示非故障开始时间被错误地标识为故障开始时间；符号 FP 表示错误开始时间被错误地标识为非故障开始时间。

$$GCS = TP \times 0 + TN \times 10 - FN \times 0.1 - FP \times 1$$

下一部分是性能说明和案例研究的讨论。

这三种常见的不平衡诊断方法，分别是投票法、平均法和学习法，将子分类器的分类结果结合起来。因此，本文将改进的gbdt、改进的rf、改进的svdd与提出的方法进行了比较。值得一提的是，并行结构有助于减少模型的方差，使最终结果更加稳定。为了验证这一理论，将训练精度和测试精度的标准差记录如下。这些结果证明了aic算法的精度和平均性能较好。总体而言，aic算法基本上保持了最小的精度标准差，即能有效地提高分类精度。可见aic算法的性能。从表中可以看出，aic算法与其他四种算法相比，具有最好的精度和稳定性。

下一部分是总结和未来的研究。1，本文结合改进的GBDT，改进的RF和改进的SVDD，提出了一种新的集成方法，用于不平衡分类应用。2，在该模型中，从数据采样和算法的角度出发，对传统的SVDD，RF和GBDT进行了优化。3，涉及改进的GBDT，改进的RF和改进的SVDD，以在数据集中建立分类模型。4，实验表明，该方法在PHM 2015数据集中具有良好的泛化性能和稳定性。5，未来的研究可以从整体学习和实际行业的角度进行。最后一部分是一些参考。好的，很高兴为您提供有关分类问题的观点。就这样。感谢您的关注！