基于模糊聚类利用 EDFA 泵电流时间序列进行退 化过程监控的异常检测

Dominic Schneider Advanced Technology Adtran 98617 Meiningen, Germany dominic.schneider@adtran.com Lutz Rapp Advanced Technology Adtran 98617 Meiningen, Germany lutz.rapp@adtran.com Christoph Ament Faculty of Applied Computer Science University of Augsburg 86159 Augsburg, Germany christoph.ament@uni-a.de

摘要—本文提出了一种基于模糊聚类的 EDFA 系统泵电流 时间序列异常检测方法。所提出的变更检测框架(CDF)战略性 地结合了熵分析(EA)和主成分分析(PCA)与模糊聚类过程 的优势。在该框架中,应用 EA 进行动态特征选择,以减少特征 空间并提高计算性能。此外,使用 PCA 从原始特征空间提取特 征,以便后续的模糊聚类过程具有泛化能力。评估了三种不同的 模糊聚类方法,更精确地说是模糊聚类算法、概率性聚类算法和 可能性聚类算法,以衡量其性能和泛化能力。因此,与商业上使用 的 EDFA 中的预定义警报相比,所提出的框架具有一种创新功 能,能够在早期阶段检测泵电流时间序列的任意操作点的变化。 此外,该方法使用实验数据进行了实现和测试。另外,所提出的 框架还使进一步应用分布式预测性维护方法于光网络成为可能。

Index Terms—机器学习,预测性维护,模糊聚类,异常检测,时间序列,退化监测

I. 介绍

铒掺杂光纤放大器(EDFAs)是当前长途光纤维传输 网络中的关键元件。由于即使是短时间的光链路故障也会 导致昂贵的传输容量损失,因此安装的 EDFAs 具有高可 靠性至关重要。这类放大器实现光数据信号放大的核心组 件包括铒掺杂光纤(EDFs)和用于激发铒离子的泵浦激光 器。由老化和退化过程引起的泵浦激光器软故障将导致整 个系统性能下降。

This work is supported by the SNS Joint Undertaken under grant agreement No. 101096120 (SEASON). Responsibility for the content of this publication is with the authors.

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. 如今, 商用的 EDFA 通常采用集成自动增益控制 (AGC), 通过控制泵电流从而稳定放大器增益。泵激光的 老化会导致在恒定泵电流下突然失效或缓慢降低泵功率, 这种现象可以通过监测随时间变化的泵电流和发出的泵功 率来检测。当获得最大输出功率所需的泵电流超过某个阈 值时, 会触发报警。

标准技术确定生成最大泵浦功率所需的泵电流是否超 过预定义的阈值。然而,当在低于其最大输出功率的情况 下操作光放大器时,老化检测并不可靠。本文提出了一种 变化检测框架来检测泵电流时间序列中的异常情况,从而 能够检测到任意工作点处泵的老化,并且可以减少提供给 评估算法的数据集大小。所提出的方法基于模糊聚类并通 过实验数据进行分析。

II. 模糊聚类程序

许多不同的方法被用于检测时间序列中的异常。最常见的方法基于随机时间序列分析、数学模型、分类、聚类等。[1] [2]。任务变得更加复杂,因为变化在一个非平稳环境中不断发展。随着时间序列的平滑缓慢发展,定义清晰集变得不可行,而模糊聚类方法在各种应用中显示出了可靠的學習和泛化行为 [3]。

聚类的目标是识别属于同一同类观测云的数据样本, 这些数据样本包含相似的数据点。对现实世界数据的这种 聚类受到噪声、缺失值等因素引起的不确定性的影响,这 些因素形成了所谓的模糊集"模糊"边界。因此,开发了 经典的模糊 c 均值 (FCM) 算法,如方程 (1) [3]

$$\begin{cases} w_j(k) = \frac{\|x(k) - c_j\|^{-2}}{\sum_{l=1}^m \|x(k) - c_l\|^{-2}} \\ c_j = \frac{\sum_{k=1}^N w_j^2(k) x(k)}{\sum_{k=1}^N w_j^2(k)} \end{cases},$$
(1)

所示,其中 w_j(k) 表示数据样本 k 的权重 j。它是通过计 算特征的 n 维向量 x(k) 从数据样本 k 到集群 j 的 n 维中心 c_j 的归一化欧几里得距离得出的。因此,w_j(k) 是衡量数 据样本与集群亲和度的一种方法。由于算法旨在为所有聚 类找到最佳中心点,需要通过将数据样本 x(k) 与 w_j(k) 的平方进行加权,并对所有数据样本进行归一化来更新初 始聚类。

在传统方法中,数据样本被分配到一个明确的聚类中。 这意味着数据样本要么属于该聚类,要么不属于。在模糊 聚类分析的情况下,数据样本是根据介于 [0,1] 中的隶属度 值被分配到聚类中的。这些方法发展成两个主要方向:概 率和可能性聚类方法。在 [4] 中,作者提出使用这两种方法 来处理时间序列中的异常检测任务。在公式 (2) 中,定义 了一个鲁棒的距离函数 *D^R(x(k),c_j)*,基于人工神经元的 标准激活函数

$$D^{R}(\boldsymbol{x}(\boldsymbol{k}), \boldsymbol{c}_{\boldsymbol{j}}) = \sum_{i=1}^{n} \beta_{\boldsymbol{i}} \ln \left[\cosh \left(\frac{\boldsymbol{x}_{\boldsymbol{i}}(\boldsymbol{k}) - \boldsymbol{c}_{\boldsymbol{j}\boldsymbol{i}}}{\beta_{\boldsymbol{i}}} \right) \right], \quad (2)$$

并为数据样本 *x*(*k*) 和聚类中心 *c_j* 之间的距离定义了一个标量度量。它利用双曲余弦与自然对数的结合来减少远离 聚类中心的异常值的放大效果。汇总所有特征 *n* 的结果为 每个数据样本到每个聚类中心生成一个度量,类似于方程 (1) 中的权重。参数 *β_i* 控制函数变化的速度,并作为进一 步正则化的超参数。

引入方程 (2) 中的度量,并在概率聚类过程 (ProbCP) 中使用它,定义了以下算法,该算法由方程 (3) 给出:

$$\begin{cases} w_{j}^{pr}(k) = \frac{\left(D^{R}(\boldsymbol{x}(\boldsymbol{k})-\boldsymbol{c}_{j})\right)^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}(D^{R}(\boldsymbol{x}(\boldsymbol{k})-\boldsymbol{c}_{j}))^{\frac{1}{1-\beta}}}\\ c_{ji}^{pr}(\boldsymbol{k}+1) = c_{ji}^{pr}(\boldsymbol{k}) + \eta(\boldsymbol{k})w_{j}^{\beta}(\boldsymbol{k}) \quad , \qquad (3)\\ \tanh\left(\frac{\boldsymbol{x}_{i}(\boldsymbol{k})-\boldsymbol{c}_{ji}(\boldsymbol{k})}{\beta_{i}}\right) \end{cases}$$

其中 $w_j^{pr}(k)$ 作为数据样本 x(k) 对现有簇中心 c_j^{pr} 的隶 属度的度量,类似于经典模糊聚类在方程 (1) 中表示的 方式,但距离是使用方程 (2) 计算的,而不是欧几里得范 数。指数 $\frac{1}{1+\beta}$ 影响距离函数的形状,并代表正则化的一个 额外超参数。与方程 (1) 相比,方程 (3) 中 $c_{ij}^{pr}(k)$ 的更新 是迭代进行的。聚类中心的先验值累加了一个梯度项,而 $\eta(k)$ 表示算法的学习率,并且加权双曲正切类似于一个梯 度值。学习率 $\eta(k)$ 可以通过迭代调整来减少算法中最终可 能出现的振荡行为。

除了使用权重外,模糊聚类过程(PossCP)还利用已 观测的数据样本构建聚类。这导致了新的度量 µ_j和由方



图 1. 数据采集设置

程(4) 定义的算法

$$\begin{cases} w_j^{pos}(k) = \left(1 + \left(\frac{D^R(\boldsymbol{x}(\boldsymbol{k}), \boldsymbol{c}_j)}{\mu_j}\right)^{\frac{1}{\beta-1}}\right)^{-1} \\ c_{ji}^{pos}(\boldsymbol{k}+1) = c_{ji}^{pos}(\boldsymbol{k}) + \eta(\boldsymbol{k})w_j^\beta(\boldsymbol{k}) \\ \tanh\left(\frac{x_i(\boldsymbol{k}) - c_{ji}(\boldsymbol{k})}{\beta_i}\right) \\ \mu_j(\boldsymbol{k}+1) = \frac{\sum_{p=1}^k w_j^\beta(\boldsymbol{p})D^R(\boldsymbol{x}(\boldsymbol{p}), \boldsymbol{c}_j(\boldsymbol{k}+1))}{\sum_{p=1}^k w_j^\beta(\boldsymbol{p})} \end{cases}$$
(4)

其中权重 $w_j^{pos}(k)$ 是通过使用方程 (2) 中的预定义距离函数和超参数 $\frac{1}{\beta-1}$ 的指数实现进行计算,用于正则化,而距离函数由 μ_j 标准化。这个标量度量观察已经呈现的数据样本,并动态调整隶属度水平取值为 0.5 的距离,这导致了 $w_j^{pos}(k) = 0.5$ 的结果。聚类中心的迭代更新类似于概率聚类过程,在公式 (3) 中表示。

III. 实验设置

所使用的数据采集设置如图 1所示,其中波分复用 (WDM) 信号由商用 EDFA 放大。该 WDM 信号包含十 个通道,在常规波长带(C带)内均匀分布。输入功率水 平设定在-35 dBm 到 1 dBm 的范围内。增益设置为从 19 dB 到 35 dB 的值。此外,最大输出功率限制为 20 dBm。 通过此设置,生成了一个包含 41 个特征的 11,886 个样本 的总数据集。

特征主要包含 EDFA 的监控值。特征分为三组:光学、 电气和温度读数。首先,在光学部分,我们有光输入功率、 光输出功率、第二级的光输入功率以及第一级的光输出功 率。其次,在电气部分,我们有关于 EDFA 各级电流和泵 浦功率的监控值以及更多的工作电压。最后,在温度部分, 我们有 EDFA 外壳的温度和泵浦激光芯片的温度。



图 2. 变化检测框架

为了检测泵电流数据时间序列中的异常,这些数据中 引入了变化。泵退化是一个研究充分的过程 [5] [6] 。由 于老化,达到所需光学泵功率所需的泵电流开始随时间增 加 [7]。实际泵电流与标称电流的比率可以用来建模这些效 应并创建用于评估算法的具体数据流。

IV. 变化检测框架

人工智能(AI)模型可能遭受维度灾难的影响。这意味着机器学习算法可以学习的特征数量依赖于算法的能力。 我们提出了一种特征选择和减少过程,而不是通过使用深 度学习架构来克服这个问题。结合聚类过程,这种方法导 致了一个变化检测框架 (CDF),如图 2所示。

数据预处理模块从数据集中移除无关的、无法解释的、 空值或重复的特征。机器学习算法对数据 [8] 的规模敏感。 因此该模块实现了一个标准缩放器来规范化数据。

特征选择 使用多种方法来测量不同特征对目标值的 影响。由于聚类任务中缺少目标值,这些算法不适合当前 问题 [9]。因此,我们提出了一种源自信息论的方法,在该 方法中,分布的信息通过其熵来衡量,如公式 (5)所示。

$$H(X) = -\sum_{k=1}^{N} \boldsymbol{p}_{\boldsymbol{k}} \ln \boldsymbol{p}_{\boldsymbol{k}} , \qquad (5)$$

其中, 熵累积了一个特征事件的加权概率。通过使用自然 对数,单位被称为 自然数。计算每个特征的熵,并选择所 有可用特征与 $H(X)_i > H_{min}$ 和 $H_{min} = 0$ 相结合的方法 是一个合适的特征选择方法,且不进行任何转换。

熵分析(EA)如图 3所示。对初始特征空间中的 41 个特征进行每个特征的熵分析并排序。将选择阈值设置为 *H_{min}* = 0,这会导致使用的特征数量减少,因为只有具有 更高熵的特征才会在进一步处理中被考虑。值得注意的是, 在应用 EA 后剩余 27 个特征。



图 3. 特征空间的熵分析

最后,特征提取使用潜在模型,即主成分分析(PCA), 来降低数据的维度。PCA 根据数据的协方差矩阵的特征向 量对数据应用线性变换。然后使用具有最高特征值的特征 向量来变换数据 [10]。PCA 应用于两个不同的数据集,即 EA 修改的数据集和原始数据集,如图 4所示。结果给出了 三个主成分,因为设置要求累积解释方差比需超过 95%。解 释方差比是一个衡量每个主成分相对于原数据库包含多少 方差的指标。第一个主成分在两个数据库中都远远超过了



图 4. 特征空间的主成分分析



图 5. 聚类过程的训练误差

其他两个主成分。通过比较两个数据集,每个主成分的偏 差相对较低,但在泛化方面会有更大的影响,如后文所示。

V. 结果与分析

A. 训练行为分析

算法在不同 CDF 设置下的学习行为如图 5所示。首 先,可以注意到收敛标准分别在 30 次迭代内达成。然而, 传统的 FCM 在所有给定的 CDF 配置的第一轮迭代中倾向 于急剧上升,这意味着在提供数据集的最佳梯度寻找上存 在问题。相比之下, ProbCP 减少了这种急剧上升,并为启 用 EA 和 EA + PCA 的 CDF 配置提供了更稳健的行为。 最后, PossCP 完全消除了急剧上升,并且对于三种聚类算 法来说表现出了最稳健的学习行为,而且与所有 CDF 配 置无关。

B. CDF的消融研究

在实验中,学习率被设定为固定值 $\eta = 10^{-3}$ 。使用 $\|\boldsymbol{W}^{t+1} - \boldsymbol{W}^t\| \le \epsilon \ \pi \ \epsilon = 10^{-4}$,创建了一个合适的停止条 件。算法的超参数被设置为 $\beta = 2 \ \pi \ \beta_i = 1$ 。为了随机方 差和权重矩阵的随机初始化的目的,运行次数等于 n = 25以平均结果。

首先,算法的性能是在原始数据上进行评估的。结果如表 I所示。均方误差(MSE)作为算法性能的度量标准,并通过训练集(*MSE*_{tr})和测试集(*MSE*_{te})进行测量。结果显示,FCM 在训练数据集上的表现较差(29.3%),与ProbCP(0.0%)和 PossCP(0.0%)相比。在测试数据集上的评估显示,没有算法能够泛化,这导致了不满意的表现。

其次,通过使用特征选择后的 EA 数据来评估算法的 性能。结果如表 I所示。EA 略微提高了训练和测试数据集 上的算法性能。然而,算法在测试数据集上的表现分别为 86.3%、78.7%和 74.3%,仍不足。 第三,我们引入 PCA 来降低数据集的维度并对特征 执行正交变换。如表 I所示, PCA 对算法在训练数据集和 测试数据集上的性能都有积极影响。PCA 使算法能够从训 练数据集中泛化,并分别将 FCM、ProbCP 和 PossCP 在 测试数据集上的性能提高了 30.3%、22.7%和 18.9%。

最后, EA 和 PCA 的组合性能如表 I所示。EA 可以 将使用 PCA 的算法性能提高最多 3.0% (基于训练数据集) 以及最多 1.4% (基于测试数据集)。这是因为主成分发生 了变化,这些主成分由转换生成,并因此产生更好的可学 习性。因此,模糊聚类算法与 EA 和 PCA 的组合是 CDF 的最佳配置。

C. 与其他方法的比较

为了评估 CDF 的性能,将其与当前最先进的方法进行了比较。在 [11] 中给出了机器学习领域可用聚类方法的 概述。因此,直接将 CDF 与其同类算法进行比较:1) K 均 值聚类,2) 层次(聚合)聚类和3)利用层次结构的平衡迭 代减少与聚类(BIRCH)。K 均值聚类最初为信号处理而发 明,并使用向量量化。在此过程中,*n*个观测值被聚集成 *k*个簇,其中最近的平均值作为簇原型。如 [12] 所示,它 已成功应用于无监督的时间序列异常检测任务中。层次聚

表 I 聚类算法的结果

CDF-Config	错误类型	模糊 C 均值	概率 CP	可能的控制点
原始	MSE_{tr}	29.3~%	0.0~%	0.0~%
	MSE_{te}	99.7~%	99.7~%	99.7~%
有效吸引力	MSE_{tr}	17.5~%	0.0~%	0.0~%
	MSE_{te}	86.3~%	78.7~%	74.3~%
主成分分析	MSE_{tr}	15.4~%	0.0~%	0.0~%
	MSE_{te}	30.3~%	22.7~%	18.9~%
PCA+EA	MSE_{tr}	12.4~%	0.0~%	0.0~%
	MSE_{te}	29.3~%	21.3~%	17.7 %

表 II 聚类算法的结果

Method	MSE_{tr}	MSE_{te}
K-Means clustering	39.87~%	45.33~%
Hierarchical clustering	39.32~%	47.53~%
BIRCH	31.67~%	34.17~%
CDF	0.0~%	17.7 %

表 III
最小变化点检测

	模糊 C 均值	概率 CP	可能 CP
I/I_0	8.1 %	$5.9 \ \%$	4.9~%

类使用链接距离、欧几里得范数或其他任何距离度量,并 递归地合并成对的簇以构建分类树 [13]。BIRCH 是一种在 线学习算法,具有高内存效率。类似于层次聚类,构造了 一个树状数据结构。最近,已证明其成功应用于检测低维 时间序列中的异常值 [14]。为了减少随机性的影响,实验 重复了 20 次,并对指标进行了平均。如表 II所示,选择 的性能度量标准为训练和测试场景中的均方误差 (MSE)。 K-Means 聚类具有较差的学习行为(39.87%),无法泛化。 由于这是一个二分类问题, 在测试表现 45.33%的情况下, 很可能出现振荡行为。层次聚类算法显示出与 K-Means 相 似的结果,训练性能为39.32%,测试性能分别为47.53%。 BIRCH 在训练(31.67%)和测试(34.17%)中的表现优于 前述的算法。值得注意的是, 层次聚类与 BIRCH 之间存在 显著的性能差异。这两个算法都构建了分层数据结构,但 相比之下,层次聚类采用自底向上的方法,而每个观测值 最初都在自己的簇中。更重要的是,所提出的 CDF 在训练 和测试中的 MSE 最小值上优于所有被比较的方法。

D. 实验性异常检测研究

CDF 的目的是检测泵电流时间序列中的异常,但训 练和测试数据集的性能指标是对打乱的数据点进行量化测 量。为了评估算法在时间序列上的性能,我们引入了变点 检测(CPD)作为度量标准。CPD 定义为使用 EA + PCA 检测到的泵电流的最小变化,如表 III所示。值 I₀ 表示名义 泵电流,而 I 表示实际泵电流。因此, I/I₀ 可以解释为漂 移。结果显示,C-Means 算法能够检测到泵电流 8.1%的漂 移。概率和可能算法分别能够检测到泵电流 5.9%和 4.9% 的漂移。因此,PossCP 能够在预定义阈值之前显著提前进 行变点检测,这些阈值通常设置为 10%,特别是在任意操 作条件下。

为了在实时系统中实现所提出的算法,对不同的数据

流进行了额外的异常识别测试。因此需要引入检查间隔的 概念。预测性维护 (PM) 框架通过参数和特征值监测系统 的健康状态或条件,在给定的时间范围内进行监控。在这 个特定时间线上的每个观察点被称为检查 [15]。在随后的 实验中,对六种不同的数据流进行了 CDF 测试。参考被表 示为完全没有退化的数据流 (DR 0%), 但存在测量噪声。 此外,将五种不同退化率的数据流 (DR 10% - DR 50%) 应 用到算法中。时间线设定包含最多150次检查。在瞬态状 态下通常会出现分类类别振荡的行为。为了对抗这种效果, 使用了移动平均滤波器。它包含了滑动窗口技术 (SWT), 这在 PM 任务中常用以捕获某一时间段内的时变信息,在 本例中为40次检查。通过利用滤波器,实现了离散分类 值到连续类别隶属度值的转换。因此定义低于 0.5 的类别 隶属度值为正常无异常(OK),高于0.5为不正常(nOK) 并包含异常。将应用数据流至 CDF 的实验结果展示在图 6 中。首先,可以注意到带有噪声的数据流 (DR 0%) 正常 操作条件下的正确分类。其次,每个含退化的数据流都经 历了从状态 OK 到 nOK 的状态转换。第三点是, 退化率越 高, CDF 越早能够检测到异常。这三个方面表明所提出的 CDF 以稳健的方式检测异常。

VI. 结论

本文提出了一种基于熵分析、主成分分析和模糊聚类 算法的新型变化检测框架,用于检测 EDFA 泵电流时间序 列中的异常情况,这些异常是由退化效应引起的。所提方 法的一个值得注意的特点是能够在一个综合框架中整合多



图 6. 异常识别测试在不同数据流上

传感器数据及任意操作条件下的数据。该基于模糊聚类的 模型以归一化和清洗后的序列数据作为主要输入,并通过 熵分析消除随机非信息性数据。此外,经过降维的数据流 将通过主成分分析进行处理,以提取方差比率最高的特征 并实现给定数据的一般化。根据漂移泵电流的实验测量结 果表明,在任意操作条件下,模糊聚类算法和新型变化检 测框架的表现具有显著的一般化能力。这使得能够在与典 型运行条件相差高达4.9%的时间序列中检测异常情况。此 外,进行的实验显示了该算法具有鲁棒性行为。特别地,可 以稳健地检测包含退化特征的数据流中的异常以及类似于 正常操作条件下的噪声数据流中的异常。在时间序列数据 中检测异常与实现预测维护的重要初始 步骤相似。降低数据维度有助于实现去中心化的实施,并 使此类系统的运行场景不会影响交通。在未来的工作中, 将考虑分析任意操作条件下异常的问题,提出 EDFA 的预 测性和诊断模型。

参考文献

- E. Gombay, "Change detection in autoregressive time series," Journal of Multivariate Analysis, vol. 99, no. 3, pp. 451–464, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0047259X07000061
- [2] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and Information Systems*, vol. 51, no. 2, pp. 339–367, 5 2017. [Online]. Available: https://doi.org/10.1007/s10115-016-0987-z
- [3] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, ser. Advanced Applications in Pattern Recognition. Springer US, 2013. [Online]. Available: https://books.google.de/books?id=z6XqBwAAQBAJ
- [4] Y. Bodyanskiy, A. Dolotov, D. Peleshko, Y. Rashkevych, and O. Vynokurova, Online Time Series Changes Detection Based on Neuro-Fuzzy Approach. Cham: Springer International Publishing, 2019, pp. 131–166. [Online]. Available: https://doi.org/10.1007/978-3-030-05645-2-5
- [5] G. Pedroza, L. Bechou, Y. Ousten, L. S. How, O. Gilard, J.-L. Goudard, and F. Laruelle, "Long term in-vacuum reliability testing of 980nm laser diode pump modules for space applications," in 2014 IEEE Aerospace Conference, 2014, pp. 1–14.
- [6] V. V. Bliznyuk, V. A. Parshin, A. G. Rzhanov, and A. E. Tarasov, "Determining the rate of degradation of a high-power laser diode from the dependence of the radiation spectrum on variations in the pumping current," *Bulletin of the Russian Academy of Sciences: Physics*, vol. 85, no. 2, pp. 184–188, 2 2021. [Online]. Available: https://doi.org/10.3103/S1062873821020064
- [7] B. Turgunov, N. Juraev, S. Toshpulatov, K. Abdullajon, and U. Iskandarov, "Researching of the degradation process of laser diodes used in optical transport networks," in 2021 International Conference on Information Science and Communications Technologies (ICISCT), 2021, pp. 1–4.
- [8] M. M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies*, vol. 9, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2227-7080/9/3/52
- [9] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 science and information conference. IEEE, 2014, pp. 372–378.
- [10] K. P. Murphy, Machine learning: a probabilistic perspective. MIT press, 2012.
- [11] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," *Ieee Access*, vol. 9, pp. 78658–78700, 2021.

- [12] N. Westa, T. Schleglb, and J. Deusea, "Unsupervised anomaly detection in unbalanced time series data from screw driving processes using k-means clustering," in 56th CIRP Conference on Manufacturing Systems. CIRP CMS, 2023.
- [13] E. K. Tokuda, C. H. Comin, and L. d. F. Costa, "Revisiting agglomerative clustering," *Physica A: Statistical mechanics and its applications*, vol. 585, p. 126433, 2022.
- [14] N. Guo and P. Wang, "Anomaly detection method of power purchase material data based on birch clustering algorithm and time series," in *Third International Conference on Advanced Algorithms and Signal Image Processing (AASIP 2023)*, vol. 12799. SPIE, 2023, pp. 113–118.
- [15] T. Tinga and R. Loendersloot, "Physical model-based prognostics and health monitoring to enable predictive maintenance," *Predictive Maintenance in Dynamic Systems: Advanced Meth*ods, Decision Support Tools and Real-World Applications, pp. 313–353, 2019.