用于无线接入网变化检测的自监督 Transformer 架构

Igor Kozlov*、Dmitriy Rivkin*、Wei-Di Chang†、Di Wu*、Xue Liu*、Gregory Dudek*
*{i.kozlov, d.rivkin, di.wu1, steve.liu, greg.dudek}@samsung.com
†w.chang1@partner.samsung.com

arXiv:2302.02025v1 [cs.LG] 3 Feb 2023

摘要——**电信无线接入网（无线接入网）是由数十万个发射设备（小区）组成的互联硬件大型集合体。这类网络会频繁发生变化，而且变化往往各不相同；网络运营商为实现最佳性能，会不断调整系统参数。此类变化的影响难以预测，而随着 5G/6G 网络的普及，预测将变得更加困难。因此，无线接入网监控对网络运营商至关重要。我们提出了一种自监督学习框架，利用自注意力和自蒸馏来完成这项任务。该框架通过检测性能测量数据中的变化来发挥作用；性能测量数据是一组时变指标，反映了小区级网络性能的多种不同测量结果。实验结果表明，在一个基于真实场景、包含约十万个时间序列的数据集上，我们的方法比最先进水平高出 4%。此外，该方法还具备可扩展性和泛化能力，能够在较少依赖专家知识的情况下，深入揭示运行模式变化的具体特征。**

索引词—**变化检测、异常检测、神经网络、蜂窝网络、无线接入网。**

I. 引言

本文提出了一种针对性能测量（PM）数据的新型变化检测系统。性能测量系统从每个小区收集数千项指标，用以量化小区性能的不同方面，包括语音质量指标、吞吐量、时延、无线电信号强度、干扰等。这些数据的体量极其庞大，目前网络运营商无法对其进行全面监控。因此，运营商更改网络配置时，往往会在未察觉的情况下对某些子系统造成负面影响。另一方面，配置更改也可能带来一些运营商未能意识到的实际积极影响，导致他们错误地回退配置更改。我们的目标是开发一种系统，用于监控PM数据并检测配置更改对系统造成的所有影响。

在这项工作中，我们通过将一种现代的自监督表征学习方法 [1], 应用于检测配置变更对 PM数据的影响，来解决这一任务。该方法最初是为处理大型无标注图像数据集而开发的。我们关注的是变化，而非异常检测。两者的区别在于，变化检测关注运行特征的稳定变化，而异常检测通常旨在检测离群值或单个极不寻常的数据点。这一区别对网络运营商而言至关重要，因为暂态异常行为按定义会自行恢复，因此运营商无需采取任何纠正措施。另一方面，检测到稳定变化意味着小区已开始以一种新的方式运行，可能需要采取纠正措施。

为满足实际应用需求，监控方案应具备以下特点：

- 可扩展性：为了实现对PM数据所报告的所有子系统进行完整监测，检测大量指标的变化至关重要。因此，变化检测器的计算复杂度应与指标数量呈线性增长。

- 通用性：无线接入网通常包含多个供应商提供的硬件，这些硬件还会持续进行软件升级，从而修改所记录的PM指标列表。该解决方案应能够以最少的专家干预适应这些变化。

为满足这些要求，我们监测单个指标，而不是考察指标的成对相关性，因为后者会导致计算资源需求呈多项式增长 [2]。通过这种方式，我们可以监测在小区级记录的所有（通常为数千个）指标，从而更深入地了解与检测到的变化相关的问题诊断信息。我们还指出，当此前共同演化的指标出现去相关时，其中至少一个指标发生了变化，因此我们预计该方法能够检测到这种变化。关于通用性要求，先前的方法（参见第II-A节）仅针对一小组由专家选定的指标运行，而这些指标可能因供应商和软件更新而有所不同。我们的解决方案考察所有 记录的PM指标，因此能够抵御此类修改，并且不需要专家干预。

由于缺乏适用于此任务的公开数据集，我们自行构建了一个。该数据集基于一款专有的系统级无线接入网模拟器，该模拟器最初开发并经过验证，用于研究网络参数优化 [3], [4]。

总而言之，本文的主要贡献如下：

- 我们提出了一个值得关注的问题：监测网络中与小区级频繁（约每周一次）异构系统参数变化相关的行为变化。
- 我们提出一种可扩展且具有良好泛化能力的自监督学习框架，利用自注意力和自

蒸馏实现指标级变化检测。

- 为了评估这项任务的性能，我们基于真实的无线接入网历史观测数据构建了一个数据集。
- 与当前最先进的非机器学习方法和机器学习方法相比，我们的方法展现出了更出色的检测能力。

本文其余部分安排如下：第II节介绍与蜂窝网络指标变化检测相关的已有工作。第III节详细介绍本文比较的三种变点检测器，包括我们提出的方法。第IV节介绍数据集的生成过程。第V节比较第III节所述方法在第IV节数据集上的性能。第VI节总结本文工作并给出结论。

II. 相关工作

本节首先讨论蜂窝网络中的异常检测研究，然后回顾现有变化检测方法的进展。

A. 蜂窝网络中的异常检测

蜂窝网络中的异常检测方法涵盖范围各异，从检测单个关键绩效指标 [5], [6], 中可观测到的异常，到使用多个关键绩效指标处理多变量情况 [7], ，再到考虑网络互连结构的网络级异常检测 [8]–[9]。本文聚焦于第一种情况。

关键绩效指标层面的异常表现为信号中的不规则现象。它们可能是短暂的（点状），也可能持续较长时间（例如，信号周期性或振幅的模式变化） [10]。这几类行为往往由不同原因引起，因此有必要引入一种专门的分类体系（“变化检测”），尤其是后者可能对网络造成更大危害，因为它意味着发生了显著转变。我们发现，现有关于蜂窝网络异常检测的研究并未明确区分这两类情况。我们在此基础上进行了改进，并将变点检测领域的研究进展拓展至无线接入网应用；下一小节将对此作简要回顾。

应用于蜂窝网络的异常检测方法涵盖范围广泛：例如自组织映射等聚类算法 [5]；例如支持向量机和自回归积分滑动平均模型等基于分类和回归的方法的集成[7]；相关性与函数分析 [11]；阈值法和基于范围的规则 [6]；生成对抗网络 [12]；以及卷积神经网络（CNN）和循环神经网络，例如长短期记忆网络 [13], [14]。 [15] 对蜂窝网络异常检测领域的现有研究进行了更深入的综述。据我们所知，我们的方法是首个将 Transformer 自注意力架构 [16] 用于蜂窝网络异常检测的方法。

无线接入网异常检测的大多数研究采用监督学习。这需要对数据集进行繁琐的人工标注，还要指定异常类别的类型 [2], [13]。 [17]中介绍了一些尝试，旨在通过采用半监督学习来应对标注方面的挑战。据我们所知，我们首次通过采用自监督学习 [18], 来应对这两项挑战；这种方法无需标签，也不需要预先了解数据中可能存在的异常类型。

B. 变化检测

关于变化检测理论与实践方面的研究已有大量成果。FLUSS 及其变体 FLOSS [19]提出了一种变点检测问题的表述：利用时间序列的矩阵轮廓检测模式一致性，将时间序列划分为由变点界定的片段，从而找出变化之间的区域，而非变化点本身。BOCPD [20]等概率方法将问题表述为估计自上一个变点以来所经历时间的概率分布，这段时间称为“运行长度”。 [21]等非参数方法基于统计双样本检验：若检验统计量超过特定阈值，则将样本判定为异常。KL-CPD [22] 扩展了 [23]中提出的核变点检测非参数方法，使其能够使用从数据中学习得到的深度核。许多基于学习和非学习的方法都遵循一种共同范式：在时间序列上滑动连续且互不重叠的半窗口，并量化左半窗口与右半窗口之间的不匹配程度，因为当两个半窗口之间的点是变点时，两者的差异应当最大。这种范式在无监督深度学习方法中尤为常见，例如 [24]。这一方法较新的实例之一是 TIRE [25]，我们将其用作机器学习基线，见第 III-B 节。

在近期一项关于变点检测方法的综述和基准测试中，作者指出：首先，没有任何一种变点检测器（CPD）能在所有数据集上表现最佳，这凸显了领域特性的重要性；其次，像 Binseg 这样的经典非机器学习变点检测器，在大量数据集上的表现始终优于某些机器学习方法，这也是我们选择它作为非机器学习基线的原因，见第 III-A 节。

III. 变化检测器

本节将更详细地介绍非机器学习（Binseg）和机器学习（TIRE）基线，以及我们的方法（TREX-DINO）。

A. Binseg

Binseg [27]是一种序列算法，会连续将时间序列分成两个窗口，并通过评估原始窗口与分割后两个窗口的代价函数之差，计算引入这次分割所获得的增益。这使我们能够将特定增益与时间序列中的每个（潜在变点）点关联起来。之后，我们在二元分类问题中将这些值用作 Binseg 变点检测算法的阈值，详情请参见第 V 节。

B. TIRE

TIRE [25] 使用一对自编码器，分别在时间序列连续窗口的时域和频域中检测变点。为此，TIRE 学习结合重构目标与时间不变相似度最小化目标的表征。学习到表征后，便将其用于基于窗口的变点检测。具体而言，算法使用两个滑动窗口（分别向后和向前查看）来判断中间的点是否为变点。滑动窗口中数据之间的相似度通过计算它们各自投影到表征空间后的余弦距离来衡量：

$$d(x,y) = 1 - \frac{\sum_{k=1}^{K} g(x)^{(k)} g(y)^{(k)}}{\sqrt{\sum_{k=1}^{K} g(x)^{(k)^2}} \sqrt{\sum_{k=1}^{K} g(y)^{(k)^2}}}, \quad (1)$$

其中， x 和 y 是对应窗口中的数据点；g 是学习得到的表征， K 是表征空间的维度。应用此滑动窗口过程后，会得到一个新的时间序列 d。为提取变点，首先对 d 进行滤波，然后对结果进行峰值检测。

C. TREX-DINO

我们提出利用自监督和 Transformer，从大型无标签数据集中学习表征。这种学习范式近期已在自然语言处理 [28] 和计算机视觉 [29]领域取得了令人瞩目的效果。我们的方法称为 TREX-DINO（使用无标签蒸馏进行时间序列表征提取），是 DINO [1], 的扩展。DINO 最初是为图像上的自监督表征学习而提出的方法。总体而言，自监督学习目标鼓励模型将同一时间序列的多种随机增强映射为相似的表征，同时鼓励模型将不同时间序列的增强映射为不同的表征。该方法采用知识蒸馏 [30] 式的学生-教师学习机制来实现这一目标。不过，我们并未使用固定的教师模型，而是定期使用学生模型权重的指数移动平均来更新教师模型权重。这实际上使教师模型实现了一种模型集成形式，类似于 Polyak-Ruppert 平均 [31], [32]。

学生网络通过交叉熵损失进行训练，使其匹配教师在所学习表征各维度上的分布。给定网络 g、输入时间序列 x、表征的大小 K 以及用于控制分布尖锐程度的温度参数 τ，表征各维度上的概率 P 定义如下：

$$P(x)^{(i)} = \frac{exp(g(x)^{(i)}/\tau)}{\sum_{k=1}^{K} exp(g(x)^{(k)}/\tau)}. \quad (2)$$

在标准的知识蒸馏设置中，优化目标可表示为

$$\min_{\theta_s} H(P_t(x), P_s(x)), \quad (3)$$

其中， θ_s 是学生网络的参数， H 是交叉熵损失。不过，为了将其调整为自监督设置，DINO 首先定义集合 V，其中包含同一时间序列的不同增强版本。关键在于， V、x_1^g 和 x_2^g 中的两项是所谓的“全局”视图。它们是原始时间序列的大幅裁剪，包含了其中的大部分内容。 x 在 V 中的其余增强版本则采用“局部”裁剪，会去除原始时间序列中更大比例的内容。在计算机视觉文献中，这种增强方式称为“多裁剪” [33]。随后，教师仅用于计算全局视图的表征，而学生则计算所有视图的表征。换言之，我们旨在优化以下优化目标：

$$\min_{\theta_s} \sum_{x \in \{x_1^g, x_2^g\}} \sum_{x' \in V, x' \neq x} H(P_t(x), P_s(x')). \quad (4)$$

教师的权重 θ_t根据学生的权重，按照以下更新规则进行更新（λ 是一个超参数）：

$$\theta_t = \lambda \theta_t + (1-\lambda)\theta_s. \quad (5)$$

通过对教师网络的输出 $g_t(x) = g_t(x) + c$进行中心化，可避免坍塌（即网络学会将所有时间序列都嵌入到完全相同的表征中）。中心 c按以下方式动态调整（m 是一个超参数， B 是批次大小）：

$$c = mc + (1-m)\frac{1}{B}\sum_{i=1}^{B} g_t(x_i). \quad (6)$$

有关该方法的更完整介绍，请参见 [1]。

DINO 最初面向图像，并使用视觉 Transformer（ViT） [29]实现。为了将该方法适配到时间序列场景，需要进行一些修改。首先，我们调整了 ViT，使其能够接收时间序列作为输入。ViT 架构会将图像拆分为图像块，每个图像块经过卷积层处理并展平，形成一系列作为 Transformer 输入的标记。TREX-DINO 将这些适用于图像的二维卷积改为更适合时间序列的一维卷积。我们保留了通道维度，因此 TREX-DINO 支持多维时间序列，不过本文重点讨论单维情况。此外，DINO 数据增强包括颜色抖动、高斯模糊和色调反转，以及多裁剪。TREX-DINO 保留了高斯模糊和多裁剪，但将颜色抖动和色调反转替换为加性高斯噪声。

学习到表征后，便按照第 III-B 节所述，使用滑动窗口法检测变化。

IV. 数据集

可以观察到，CPD 在不同来源的时间序列上的性能存在差异 [26], ，这是由各领域特有的数据特性所致（例如季节性、记录的周期和长度、变化类型及其发生频率）。

这正是领域特定任务数据对于评估所提方法的性能至关重要的主要原因。

本文使用了一款专有的系统级仿真器（SLS） [3], [4],，该仿真器专为研究系统参数变化对蜂窝网络性能的影响而开发，并基于真实世界的无线接入网数据。它对特定站点配置架构下的无线接入网业务进行建模。我们采用了图 1 中示意展示的架构。每个

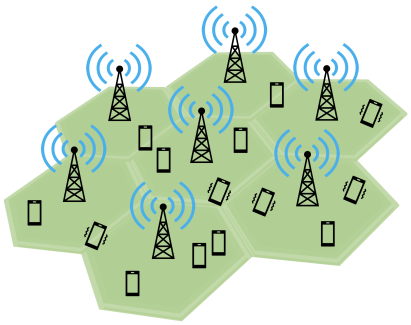


图 1. SLS 环境示意图，该环境由 7 个站点组成，为若干活跃和空闲的用户设备（UE）提供服务。

站点均在 3 个互不重叠的扇区中运行，每个扇区使用 4 个频段。用户设备（UE）最初均匀分布在所有站点，并以恒定速度随机移动，或停留在原地。它们通过随机下载文件，为网络贡献业务负载。该过程经过建模，以复现历史上观察到的无线接入网数据场景，用户设备的行为参数也会相应地随一天中的时间变化 [3], [4]。向用户设备传输数据的小区也预设了符合实际情况的系统参数，但这些参数不会随时间自动变化。

本文的主要目标是通过监控PM数据来检测网络运行中的变化。每个小区记录了28种不同的时间序列，其中包括网络运营商密切监控的关键绩效指标（KPI），例如下行吞吐量；也包括文献中常用于对无线接入网性能建模的指标，例如物理资源块使用量和活跃用户设备数量。上述所有时间序列均由专家精心挑选，并且常借助领域知识进行转换，例如，吞吐量是无线链路数据与时延的比值。正如第I节所述，我们寻求一种可泛化的无线接入网监控方法，因此还在分析中纳入了大量底层子系统指标，例如成功切换的次数，并一视同仁地对待所有这些指标。

除了上述由历史无线接入网数据中的用户设备参数所带来的变化外，我们的数据集还包含另外两个自由度。首先是低频发生的系统参数变化。这些变化会影响描述小区子系统的 SLS 参数。这样做是为了模拟由内生原因（例如配置管理（CM）变更）和外生原因（例如硬件故障）导致的无线接入网性能提升或下降；这些原因正是网络运营商所关注的。其次是伪随机变化。这些变化会影响用户设备动态，有助于验证网络运行的稳定性，并为数据集增添真实的方差。

我们的数据集与以往SLS应用之间的主要区别如下：

- 我们并不关注单个用户设备所能提供的动态信息，因为出于隐私和技术方面的原因，更容易获取按用户设备聚合的PM数据，并以此为基础进行变化检测。
- 为准确建模多样化的流量场景，SLS 使用来自现场部署的多项历史网络观测数据，这些数据的记录长度有限（最长为一周）。我们提出采用的机器学习方法，尤其是深度学习方法（Transformer），无需人工特征工程即可提取特征。为了高效地实现这一点，这些方法需要访问大型数据语料库，例如，需要足够的历史信息来检测持续存在的运行模式及其间的变化。为此，我们通过拼接 SLS 数据的一周输出，生成数据集样本。

下面详细介绍我们构建数据集样本的流程。我们针对图 1 所述的网络，使用 5 种不同系统参数组合和 10 个不同随机种子运行一周的 SLS 仿真。然后，我们选取一对不同的系统参数变化 κ 和随机种子 ρ，将一周的时间序列 $\tilde{x}$ 拼接成 4 个为期四周的样本 $\hat{x}$:

$$\hat{x}_t^m(\rho_{r_1}, \kappa_{k_1}) = \begin{cases} \tilde{x}_t^m(\rho_{r_1}, \kappa_{k_1}) & \text{if } t \in [0, 168), \\ \tilde{x}_{t-168}^m(\rho_{r_2}, \kappa_{k_1}) & \text{if } t \in [168, 336), \\ \tilde{x}_{t-336}^m(\rho_{r_1}, \kappa_{k_2}) & \text{if } t \in [336, 504), \\ \tilde{x}_{t-504}^m(\rho_{r_2}, \kappa_{k_2}) & \text{if } t \in [504, 672), \end{cases} \tag{7}$$

其中，$\tilde{x}_t^m(\rho_r, \kappa_k)$ 表示使用随机种子 r 且系统参数设定为 k 的一周仿真中，时间 t （以小时为单位）的指标 m 记录；下标表示订阅变量的值是否相等。

最后，我们按照标准流程对构建的时间序列 $\hat{x}$ 进行预处理：采用对数变换、去均值并缩放至单位方差，以平滑数据并使其适用于神经网络，从而得到数据集 [34], [35], x。图 2 展示了处理后的时间序列 x 示例。图 2 中这类时间序列模式变化没有“骤降” [2], ，但明显表明系统参数的变化导致运行特征发生偏移，而网络运营商希望了解这种变化。我们将其纳入其中，以突出在电信领域尤为相关的新型领域特定异常。

为了创建标签测试集以评估 CPD 的性能，同时考虑到系统参数切换可能不会影响每项指标，我们利用对数据集构建方式的了解（变点的位置以及子片段以周为单位的特性），通过逐周统计量对样本进行聚类，并且仅将来自那些

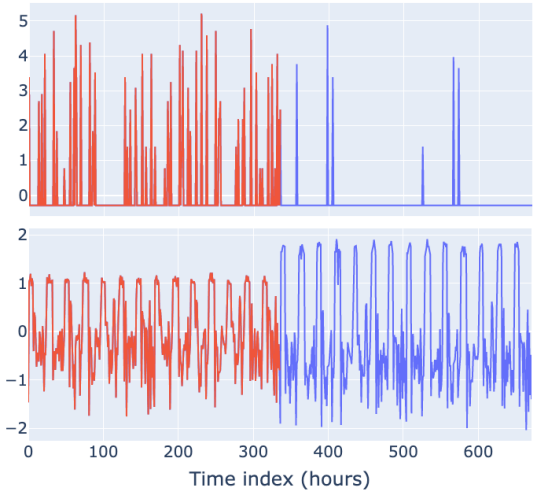


图 2. 使用 SLS 创建的离散（上）和连续（下）时间序列数据示例。不同颜色表示不同的系统参数集。

明确存在或不存在运行模式变化的聚类中的样本保留为验证样本，最终得到 78455 个没有模式变化的样本和 14301 个有模式变化的样本。

V. 性能评估

本节使用第 IV 节介绍的数据集，评估第 III 节所述变点检测算法的性能。为简化比较，我们在二元分类任务上评估性能：竞争算法需要判断时间序列中是否存在变点。

我们使用 Truong 等人提供的 Binseg 实现[36]，并采用径向基函数代价函数 [37]。

TIRE 和 TREX-DINO 均使用随机选取的数据集一半进行训练，并与 Binseg 一起在另一半数据上进行测试。

TIRE 和 TREX-DINO 训练所用的超参数值基于 [1], [25]。具体而言，对于 TIRE，我们在时间序列和频域中均使用了 3 个并行自编码器；对于 TREX-DINO 的 DINO 子部分，我们使用了 12 个自注意力头，嵌入维度大小为 768，高斯噪声标准差为 0.3，并使用核标准差在 [0.1 至 2.] 范围内均匀分布的高斯模糊，平均应用于一半的数据增强调用。

我们根据领域特性手动设置的唯一参数是半窗口大小，设为 168（等于24×7）。实际上，这使系统倾向于进行逐周比较，这也是电信运营商评估时的常见做法。

所有对比模型中还存在一个重要的超参数，即检测阈值：预测值高于该阈值的时间序列被认为存在变点。我们没有使用验证集为每种方法分别调节该阈值，而是选择在各方法性能最佳时进行数值比较，并在整个阈值取值范围内进行可视化比较：由于某些性能曲线的峰值较为尖锐，这种方式具有更好的稳健性。

由于算法的性质不同，因而得到的阈值也不同，我们将曲线统一归一化到 $[0, 1]$ 范围后进行绘制。

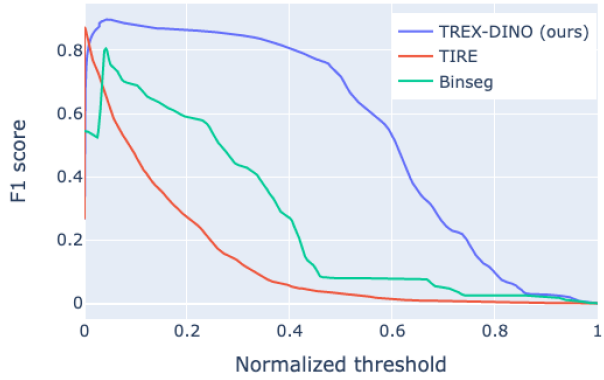


图 3. CPD 的性能随归一化检测阈值变化的差异比较。

三种方法的 F1 曲线如图 3 所示。各方法取得的最高 F1 分数和精确率-召回率曲线下面积（精确率-召回率曲线下面积）如表 I 所示。TREX-DINO 不仅在三种方法中取得了最高的最高 F1 分数，而且在更宽的归一化阈值范围内都能保持较高的 F1 分数，这表明它对阈值的选择不太敏感。其优异的精确率-召回率曲线下面积得分也体现了这一点。

表 I 使用最高 F1 分数和精确率-召回率曲线下面积对 CPD 进行汇总比较。

指标	Binseg	TIRE	TREX-DINO（我们的方法）
F1 分数	0.80	0.87 (+9%)	**0.90 (+13%)**
精确率-召回率曲线下面积	0.82	0.91 (+11%)	**0.97 (+18%)**
Rank	基线	更优	Best

VI. 结论

本文讨论了对无线接入网进行监控以实现频繁变化检测的重要性，并提出了一种新颖的变点检测器来解决这一问题。由于缺乏专门用于此任务的公开基准数据集，我们开发了一个数据集，以复现无线接入网运营商日常面临的各种挑战。实验表明，与现有方法相比，我们的方法具有更优的性能。我们相信，我们的框架已在计算机视觉和自然语言处理中得到验证，并基于具有自蒸馏机制的自监督 Transformer 架构，因此将在实际无线接入网应用中继续展现良好的性能特征。

为了将数十万台无线接入网设备在指标层面检测到的变化呈现给人类网络运营商，以便采取切实可行的行动，需要将这些变化聚合为具有共同根因的有意义事件。我们将应用于这一方向的研究成果留待日后发表。然而，任何

此类聚合方案都必须以准确的指标层面变化检测为基础，而这正是我们的方法 TREX-DINO 最擅长解决的问题。

致谢

作者感谢 Ekram Hossain、Perouz Taslakian、Amal Feriani 和 Bassem Maraach 提供的有益讨论。

参考文献

[1] M. Caron、H. Touvron、I. Misra、H. J´egou、J. Mairal、P. Bojanowski 和 A. Joulin，《自监督视觉 Transformer 中涌现的特性》，2021［ ］ 2104 14294 。在线。网址：https://arxiv.org/abs/。[2] J. Wu、P. P. C. Lee、Q. Li、L. Pan 和 J. Zhang，《Cellpad：通过回归分析检测蜂窝网络中的性能异常》，载于*2018 IFIP*网络会议（*IFIP Networking*）及研讨会，2018年，第1–9页。[3] J. Kang、X. Chen、D. Wu、Y. T. Xu、X. Liu、G. Dudek、T. Lee 和 I. Park，《用于混合通信负载均衡的分层策略学习》，载于*ICC 2021 - IEEE*国际通信会议，2021年，第1–6页。[4] A. Feriani、D. Wu、Y. T. Xu、J. Li、S. Jang、E. Hossain、X. Liu 和 G. Dudek，《多频段下行蜂窝网络的多目标负载均衡：一种元强化学习方法》，*IEEE*通信领域精选期刊，第40卷，第9期，第2614–2629页，2022年。[5] G. Barreto、J. Mota、L. Souza、R. Frota 和 L. Aguayo，《通过竞争神经模型对3G蜂窝网络进行状态监测》，*IEEE*神经网络汇刊，第16卷，1075，2005 5，1064 no. pp. –。[6] X. Qin、S. Tang、X. Chen、D. Miao 和 G. Wei，《蜂窝网络中的Sqoe kqis异常检测：采用沙漏聚类的快速在线检测框架》，*China Communications*，第15卷，第10期，第25–37页，2018年。[7] G. F. Ciocarlie、U. Lindqvist、S. Nov´aczki 和 H. Sanneck，《使用集成方法检测蜂窝网络中的异常》，载于第九届国际网络与服务管理会议（*CNSM 2013*）论文集，2013年，第171–174页。[8] C. Chaparro 和 W. Eberle，《使用基于图的方法检测移动电信网络中的异常》，载于第二十八届国际*FLAIRS*会议，2015年。[9] C. Yang、H. Song、M. Tang、L. Danon 和 Y. Vigfusson，《用于传染病监测的手机通话详单动态网络异常建模》，载于第*28*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，2022年，第4733–4742页。[10] K.-H. Lai、D. Zha、J. Xu、Y. Zhao、G. Wang 和 X. Hu，《重新审视时间序列离群点检测：定义与基准数据集》，载于神经信息处理系统数据集与基准专题论文集，J. Vanschoren 和 S. Yeung 编，第1卷，2021年。[11] P. Mu˜noz、R. Barco、I. Serrano 和 A. G´omez-Andrades，《基于相关性的时间序列分析：检测SON中的小区退化》，*IEEE*通信快报，第20卷，第2期，第396–399页，2016年。[12] J. Huang、E. Kurniawan 和 S. Sun，《利用生成对抗网络和时间序列分解进行蜂窝网络KPI异常检测》，载于*ICC 2022 - IEEE*国际通信会议，2022年，第4074–4079页。[13] M. R. Tanhatalab、H. Yousefi、H. M. Hosseini、M. M. Bonab、V. Fakhar-ian 和 H. Abarghouei，《Deep RAN：利用循环卷积神经网络检测实时蜂窝网络异常的可扩展数据驱动平台》，载于*2020*年*IEEE*第*18*届应用机器智能与信息学世界研讨会（*SAMI*），2020年，第269–274页。[14] S. M. A. Al Mamun 和 M. Beyaz，《用于蜂窝移动网络异常检测的长短期记忆循环神经网络（RNN）》，载于网络机器学习，´ E. Renault、P. M¨uhlethaler 和 S. Boumerdassi 编。Cham：Springer International Publishing，2019年，第222–237页。[15] B. Li、S. Zhao、R. Zhang、Q. Shi 和 K. Yang，《利用大数据分析检测蜂窝网络异常》，*IET*通信，第13卷，第20期，第3351–3359页，2019年。 [在线]。网址：https: //ietresearch.onlinelibrary.wiley.com/doi/abs/ 10 1049 2019 0765 。/ [16] iet-com。。A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、L. Kaiser 和 I. Polosukhin，《Attention is All You Need》，2017年。[在线]。网址：https://arxiv.org/abs/1706.03762 [17] B. Hussain、Q. Du 和 P. Ren，《基于半监督学习的移动无线网络大数据驱动异常检测》，China Communications，第15卷，第4期，第41–57页，2018年。[18] J.-B. Grill、F. Strub、F. Altch´e、C. Tallec、P. H. Richemond、E. Buchatskaya、C. Doersch、B. A. Pires、Z. D. Guo、M. G. Azar、B. Piot、K. Kavukcuoglu、R. Munos 和 M. Valko，《Bootstrap Your Own Latent：一种自监督学习新方法》，2020年。 ［ ］ 2006 07733 在线。网址：https://arxiv.org/abs/。[19] S. Gharghabi、Y. Ding、C.-C. M. Yeh、K. Kamgar、L. Ulanova 和 E. Keogh，《Matrix Profile VIII：领域无关的在线语义分割，达到超人类性能水平》，载于2017年IEEE国际数据挖掘会议（ICDM）。IEEE，2017年，第117–126页。[20] R. P. Adams 和 D. J. C. MacKay，《贝叶斯在线变点检测》，2007［ ］ 0710 3742 。在线。网址：https://arxiv.org/abs/。[21] D. S. Matteson 和 N. A. James，《一种用于多元数据多变点分析的非参数方法》，美国统计学会会刊，第109卷，第505期，第334–345页，2014年。[22] W. Chang、C. Li、Y. Yang 和 B. P´oczos，《利用辅助深度生成模型进行核变点检测》，载于第七届国际学习表征会议（ICLR 2019），美国路易斯安那州新奥尔良，2019年5月6日至9日。OpenReview.net，2019年。 [在线]。网址：https://openreview.net/forum?id = 1 7 r GbfhRqF。[23] Z. Harchaoui、E. Moulines 和 F. Bach，《核变点分析》，载于《神经信息处理系统进展》，D. Koller、D. Schuurmans、Y. Bengio 和 L. Bottou 编，第21卷。Curran Associates, Inc.，2008年。[24] W.-H. Lee、J. Ortiz、B. Ko 和 R. Lee，《通过自动特征学习进行时间序列分割》，2018年。 [在线]。网址：https://arxiv.org/abs/ 1801 05394 。T. De Ryck、M. De Vos 和 A. Bertrand，[25]《使用具有时不变表示的自编码器进行时间序列数据变点检测》，IEEE信号处理汇刊，第69卷，第3513–3524页，2021年。[26] G. J. J. v. d. Burg 和 C. K. I. Williams，《变点检测算法评估》，2020年。 [在线]。网址：https://arxiv.org/abs/ 2003 06222 。A. Scott 和 M. Knott，[27]《方差分析中均值分组的聚类分析方法》，Biometrics，第30卷，第507页，1974年。[28] J. Devlin、M.-W. Chang、K. Lee 和 K. Toutanova，《BERT：用于语言理解的深度双向变换器预训练》，2018年。［ ］ 1810 04805 在线。网址：https://arxiv.org/abs/。[29] A. Dosovitskiy、L. Beyer、A. Kolesnikov、D. Weissenborn、X. Zhai、T. Unterthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly、J. Uszkoreit 和 N. Houlsby，《一张图像胜过16×16个词：用于大规模图像识别的Transformer》，2020年。 [在线]。网址：https://arxiv.org/abs/ 2010 11929 。G. Hinton、O. Vinyals 和 J. Dean，[30]《蒸馏神经网络中的知识》，2015［ ］ 1503 02531 。在线。网址：https://arxiv.org/abs/。[31] B. T. Polyak 和 A. B. Juditsky，《通过平均化加速随机逼近》，SIAM控制与优化杂志，第30卷，第4期，第838–855页，1992年。[32] D. Ruppert，《从缓慢收敛的Robbins-Monro过程高效估计》，康奈尔大学运筹学与工业工程系，技术报告，1988年。[33] M. Caron、I. Misra、J. Mairal、P. Goyal、P. Bojanowski 和 A. Joulin，《通过对比聚类分配进行视觉特征的无监督学习》，2020年。 [在线]。网址：https://arxiv.org/abs/ 2006.09882[34] F. Pedregosa、G. Varoquaux、A. Gramfort、V. Michel、B. Thirion、O. Grisel、M. Blondel、P. Prettenhofer、R. Weiss、V. Dubourg、J. Vander-plas、A. Passos、D. Cournapeau、M. Brucher、M. Perrot 和 E. Duch-esnay，《Scikit-learn：Python中的机器学习》，Journal of Machine Learning Research，第12卷，第2825–2830页，2011年。[35] A. Paszke、S. Gross、S. Chintala、G. Chanan、E. Yang、Z. DeVito、Z. Lin、A. Desmaison、L. Antiga 和 A. Lerer，《PyTorch中的自动微分》，2017年。[36] C. Truong、L. Oudre 和 N. Vayatis，《变点检测方法综述》，CoRR，第abs/1801.00718卷，2018年。 [在线]。网址：http://arxiv.org/abs/ 1801 00718 。S. Arlot、A. Celisse 和 Z. Harchaoui，[37]《一种通过模型选择实现的核多变点算法》，2012年。 [在线]。网址：https://arxiv.org/abs/1202.3878