Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy

摘要

Anomaly Transformer 是最近在时序异常检测领域取得进展的一种基于 Transformer 的无监督方法。该方法的创新在于提出了异常点关联的局部集中性可以作为区分正常和异常时间点的新的判别准则。具体来说,Anomaly Transformer 通过自定义的异常注意力机制同时建模时间点的局部和全局关联性, 并定义了关联差异的概念, 将先验关联和序列关联的对称 KL 散度作为判别标准。本文在 Anomaly Transformer 的基础上, 进行了复现与扩展。在复现方面, 本文重新实现了 Anomaly Transformer 模型, 并在 SMD、MSL、SMAP、SWAT 和 PSM 数据集上进行了训练与评估, 结果与原文基本一致, 验证了模型的有效性。在扩展方面, 本文进行了两点创新: 第一, 应用 Optuna 框架对模型超参数进行自动化搜索与优化, 特别是 anomaly ratio 的优化, 较原文提高了模型的适应性。第二, 将 Anomaly Transformer 应用于网络入侵检测数据集, 初步探索了该方法在网络异常检测方面的潜力。综上, 本文对 Anomaly Transformer 进行了复现与拓展, 验证了其在时序异常检测任务上的有效性, 为后续研究提供了参考与借鉴。本文的超参数自动优化与网络安全应用为原文进行了有益的补充。

关键词: 时序异常检测; 网络异常检测; 无监督学习; Transformer; Optuna

1 引言

在过去的几年里，时间序列中无监督检测异常点的问题一直备受关注。无监督的异常检测是一个复杂的挑战，需要模型能够无监督地推导出一个足够具有判别性的判据去区分异常。以往的方法主要通过学习逐点表示或成对关联来解决这个问题，然而，这两者都不足以理解复杂的动态。最近，Transformer 在统一建模逐点表示和成对关联方面表现出色，与之相关的，本报告关注的论文是由 Xu 等人荣获国际表征学习大会 ICLR 2022 Spotlight 的一篇名为 “Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy” [1] 的论文，该论文提出了一种新的基于 Transformer 的时序异常检测方法。本文将复现该文章方法与评估实验，并额外增加该方法在网络异常检测，尤其是网络入侵检测领域的应用探索。

1.1 选题背景

在计算机网络领域，时间序列异常检测与网络异常检测密切相关，特别是在涉及网络入侵检测的情境下。网络流量时序异常检测对于确保网络的安全性至关重要，然而在这个领域

存在着一系列挑战。传统的方法面临着由网络攻击不断演变、硬件、软件和网络拓扑的不断进步而带来的困难。

网络入侵已经呈现出随着硬件、软件和网络拓扑的不断发展而日益复杂和高级化的趋势。因此，入侵检测系统在防御网络免受恶意网络攻击方面变得尤为重要。最近的研究表明，深度学习技术在网络安全领域表现出色，尤其是在处理高维度和非线性属性的数据方面 [2]。然而，针对网络异常检测任务，由于数据本身的类别不平衡问题，传统的机器学习方法存在明显的偏见问题 [3]。

1.2 论文背景

实际系统中的运行通常以连续的方式生成多个连续的测量，例如工业设备、空间探测等。从这些大规模系统监测数据中发现故障变成了在时间序列中检测异常时间点的任务，这对于确保系统安全和避免财务损失至关重要。然而，异常通常是罕见的，而且常常被大量正常数据掩盖，导致使用监督学习进行数据标记变得困难且昂贵。因此，在无监督设置下进行时间序列异常检测便尤为重要。

无监督时间序列异常检测在实际中面临着巨大的挑战。模型需要通过无监督任务学习复杂的时间动态，并推导出一个可区分的准则，以从众多正常时间点中检测出罕见的异常时间点。传统的异常检测方法提供了许多无监督的范例，如密度估计方法和基于聚类的方法。然而，这些方法往往难以推广到真实场景，并且无法处理时间信息。

近年来，基于深度学习的方法在时间序列异常检测领域取得了显著的成果。这些方法主要分为两类：一类通过设计良好的循环神经网络学习时序数据的逐点表示，通过重建或自回归任务进行自监督学习；另一类基于显式关联建模来检测异常。然而，这些方法在捕捉细粒度关联或推广到复杂时间模式方面仍存在一定的挑战。同时，近来 Transformer 在多个领域取得了显著的进展，包括自然语言处理、机器视觉和时间序列分析。Transformer 的成功有赖于它在全局表示和长距离关系的统一建模方面的出色表现，这为时序异常检测提供了新的方向。在这样的背景下，本报告复现的论文提出了一种新的异常检测方法，将 Transformer 引入时间序列异常检测的无监督环境中。

1.3 选题意义

本报告关注了网络异常检测，特别是网络入侵检测方面存在的挑战与意义。在计算机网络领域，入侵检测与异常检测密切相关，占据着大量的研究工作。机器学习技术的发展与进步虽然为入侵检测提供了重要的解决思路，然而，依旧存在着许多问题有待解决。

因此，通过对 Anomaly Transformer 进行复现并探索其在网络流量等领域的应用，可以深入了解其在复杂网络数据上的检测能力，为网络入侵检测领域的研究和应用提供有价值的参考。

2 相关工作

2.1 基于机器学习的网络异常检测研究

许多传统的机器学习算法，例如支持向量机 [4]、决策树 [5]、朴素贝叶斯 [6]、随机森林 [7] 等方法被用于网络异常检测。然而，这些机器学习方法存在着限制，只有在异常与正常比例均衡的网络流量数据上，它们才能发挥应有的作用。而在比例不均衡的数据集上，大部分上述机器学习算法的标准版本受网络流量数据中的大量主要类别的数量影响，会得出有偏的结果，即算法会偏好相对占大多数的类而非相对占少数的类。

近年来，深度学习在计算机视觉、自然语言处理等领域取得令人瞩目的成就。深度学习也被应用到网络异常检测中 [2, 8]。与传统机器学习方法不同，深度学习方法可以在比例不均衡的网络流量数据上学习并产生相对无偏的结果。此外，深度学习还具备处理高维度、非线性属性的能力。

2.2 无监督时序异常检测

无监督时间序列异常检测是一个重要的现实问题，已经得到广泛研究。根据异常确定标准，主要的方法包括密度估计、基于聚类、基于重构和基于自回归的方法。

密度估计方法：经典的密度估计方法包括局部离群因子（LOF）[9] 和连接性离群因子（COF）[10]，它们分别计算局部密度和局部连接性以确定异常值。此外，一些方法如 DAGMM [11] 和 MPPCACD [12] 采用了高斯混合模型来估计表示的密度。

基于聚类的方法：这类方法通常将异常分数形式化为与聚类中心的距离。SVDD [13] 和 Deep SVDD [14] 是将表示从正常数据聚集到紧凑簇的方法。另一个方法 THOC [15] 通过多层次聚类机制融合来自中间层的多尺度时间特征，并通过多层距离来检测异常。ITAD [16] 则在分解的张量上进行聚类。

基于重构的模型：这些模型尝试通过重构误差来检测异常。比如，LSTM-VAE [17] 模型采用 LSTM 进行时间建模，利用变分自动编码器（VAE）进行重构。OmniAnomaly [18] 进一步扩展了 LSTM-VAE 模型，引入了一个归一化流，并使用重构概率进行检测。InterFusion [19] 则将骨干模型改进为分层 VAE，以同时建模多个时间序列之间的相互依赖关系。生成对抗网络（GANs）[20] 也被用于重构异常检测，并作为对抗性正则化的一部分。

基于自回归的模型：这类模型通过预测误差来检测异常。VAR 扩展了 ARIMA [21]，并基于滞后相关性进行未来预测。LSTMs [22] 也被用来替代自回归模型。

本报告所关注论文方法是一种基于重构的方法，其特点是，基于一种新的基于关联的判别标准来实现对时序数据关联信息建模学习，并用于时序异常点检测。

3 本文方法

3.1 本文方法概述

无监督地对时间序列进行异常检测存在挑战，因为这需要模型自行学习得到区分异常的标准。先前工作主要有逐点表示学习和显式关联建模方法两种，前者侧重于自监督学习序列

表示，但由于异常点罕见性而不足以捕捉复杂的异常模式；后者侧重于直接建模时间点之间的关联性，然而其在学习信息丰富表示和细粒度关联方面存在挑战。

核心观察：在无监督时间序列异常检测中，异常点的关联主要集中在它们相邻的时间点上，而不是整个时间序列。这种关联的局部集中性提供了一种天然的区分正常和异常时间点的标准。

方法思路：基于核心观察，文章利用高斯核函数的函数性质来体现相邻时间点的局部关联性（命名为 Prior-Association；两两时间点下标之差的绝对值输入高斯分布函数，因此距离近的时间点得到更大的权重，反之更小，因此可以建模局部关联性），并用自注意力机制变体来建模全局关联性（命名为 Series-Association）。两者的计算均被放入了多层结构的 Anomaly Transformer 中，输出结果用 KL 散度来计算两种关联性的相似度，用于表征逐点关联差异（Association Discrepancy），并采用了 MINIMAX 优化策略，以增强关联差异的可区分性。根据核心观察，异常点的关联差异应该要比正常时间点更小。最终将归一化关联差异纳入 AnomalyScore 来作为逐点异常判定标准。Anomaly Transformer 的架构示意图如图 1所示。

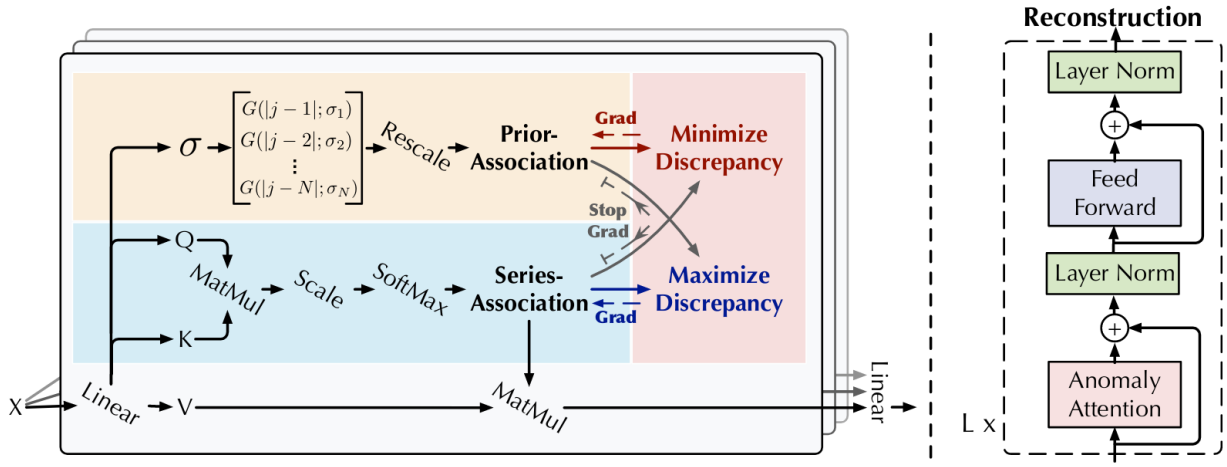


图 1. Anomaly Transformer 架构示意图

3.2 异常注意力 Anomaly Attention

由于经典的点积缩放自注意力机制 [23] 无法同时建模局部关联性和全局关联性，因此文章设计了一种双分支结构的异常注意力。对于序列关联，利用点积缩放注意力建模时间点与整个序列的关系；而对于先验关联，则基于异常点附近更有可能存在异常点的先验知识，利用高斯核函数的分布特性和连续性建模不同间隔时间点的局部时间关联。根据核心观察，两种关联性在异常点和正常点间应具有可辨别的差异。异常注意力沿用了多头设置，共有 h 个头，在第 l 层的公式如下：

$$\text{初始化：}\ \mathcal{Q},\mathcal{K},\mathcal{V},\sigma = \mathcal{X}^{l-1}W_{\mathcal{Q}}^{l},\mathcal{X}^{l-1}W_{\mathcal{K}}^{l},\mathcal{X}^{l-1}W_{\mathcal{V}}^{l},\mathcal{X}^{l-1}W_{\sigma}^{l}$$

$$\text{先验关联：}\ \mathcal{P}^{l} = \text{Rescale}\left(\left[\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{|j-i|^2}{2\sigma_i^2}\right)\right]_{i,j\in\{1,\cdots,N\}}\right)$$

$$\text{序列关联：}\ \mathcal{S}^l = \text{Softmax}\left(\frac{\mathcal{Q}\mathcal{K}^{\mathrm{T}}}{\sqrt{d_{\text{model}}}}\right)$$

$$\text{重建任务输出：}\ \widehat{\mathcal{Z}}^l = \mathcal{S}^l\mathcal{V}$$

3.3 关联差异 Association Discrepancy

关联差异被定义为先验关联和序列关联两个离散分布之间的对称 KL 散度，即两个分布的差异性，公式如下：

$$\text{AssDis}(\mathcal{P},\mathcal{S};\mathcal{X}) = \left[\frac{1}{L}\sum_{l=1}^{L}\left(\text{KL}(\mathcal{P}_{i,:}^l\|\mathcal{S}_{i,:}^l) + \text{KL}(\mathcal{S}_{i,:}^l\|\mathcal{P}_{i,:}^l)\right)\right]_{i=1,\cdots,N}$$

异常点的序列关联和先验关联两个分布的差异大，而正常点差异小。KL 散度越小两个概率分布越匹配，即越可能是正常点。基于此，作者设计了基于关联的异常判别标准，并用到重建任务：

$$\text{AnomalyScore}(\mathcal{X}) = \text{Softmax}\left(-\text{AssDis}(\mathcal{P},\mathcal{S};\mathcal{X})\right) \odot \left[\|\mathcal{X}_{i,:} - \widehat{\mathcal{X}}_{i,:}\|_2^2\right]_{i=1,\cdots,N}$$

3.4 极化策略 Minimax Streategy

极化策略的核心思想在于额外损失函数用于引导序列关联更多关注非临近区域，从而使异常点更难被重建，增加正常点与异常点差异。文章设置了两个阶段来进行极化，在最小化阶段，让先验关联去近似序列关联；在最大化阶段，优化序列关联以扩大关联差异。极化策略下两种关联的学习过程如图 2所示。两个阶段的囊括重建损失后的损失函数分别为：

$$\text{最小化阶段：}\ \mathcal{L}_{\text{Total}}(\widehat{\mathcal{X}},\mathcal{P},\mathcal{S}_{\text{detach}},-\lambda;\mathcal{X})$$
$$\text{最大化阶段：}\ \mathcal{L}_{\text{Total}}(\widehat{\mathcal{X}},\mathcal{P}_{\text{detach}},\mathcal{S},\lambda;\mathcal{X}),$$

其中：

$$\mathcal{L}_{\text{Total}}(\widehat{\mathcal{X}},\mathcal{P},\mathcal{S},\lambda;\mathcal{X}) = \|\mathcal{X}-\widehat{\mathcal{X}}\|_{\text{F}}^2 - \lambda \times \|\text{AssDis}(\mathcal{P},\mathcal{S};\mathcal{X})\|_1$$

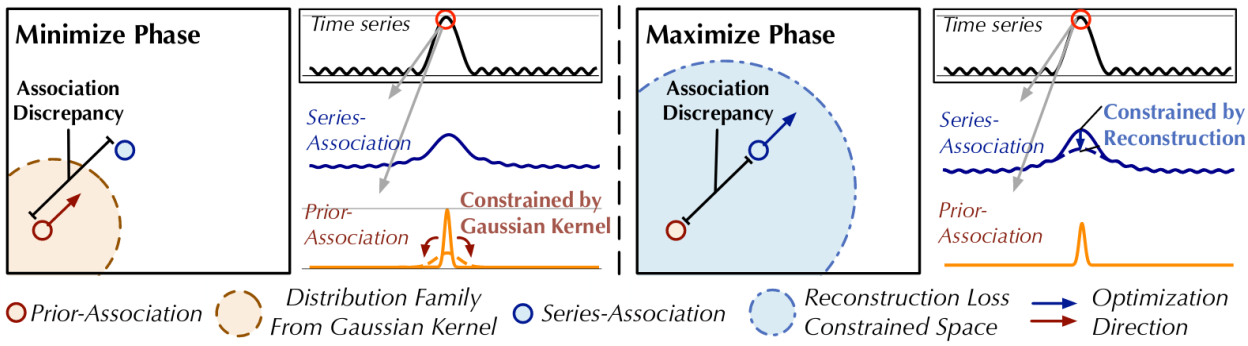


图 2. 极化策略中两种关联的学习过程示意图

4 复现细节

4.1 与已有开源代码对比

原文公开了论文方法的实现与评估 python 代码在 Github：https://github.com/thuml/Anomaly-Transformer，其中评估部分不完全。本次复现新增内容主要为以下三点：

- 基于 Optuna [24] 实现了对原代码训练与评估的自动化并行超参数多目标搜索优化。
- 增加 SWaT 数据集的评估代码
- 增加对 NSLKDD 数据集 [25] 的训练评估代码

4.2 实验环境搭建

克隆 Anomaly Transformer 的代码仓库到本地，在已经安装 conda 22.9.0，并用 conda 创建了 python 3.6 虚拟环境（环境命名为 Anomaly-Transformer）的前提下，使用 pip 安装 pytorch 1.8.0，scikit-learn 和 pandas 这些必备运行库。接着需要准备评估数据集，其中 SMD、PSM、SMAP 和 MSL 四个数据集已由作者通过 Google Drive 分享在https://drive.google.com/drive/folders/1gisthCoE-RrKJ0j3KPV7xiibhHWT9qRm?usp=sharing。但 SWaT 数据集由于使用协议限制而不能与他人共享，需要自行向 iTrust 申请：https://docs.google.com/forms/d/e/1FAIpQLSdwOIR-LuFnSu5cIAzun5OQtWXcsOhmC7NtTbb-LBI1MyOcug/viewform?usp=sf_link，成功申请到后，需要进入到 SWAT/SWaT.A1&A2_Dec 2015/Physical/文件夹下，并将 SWaT_dataset_Attack_v0.xlsx 作测试集，将 SWaT_dataset_Normal_v1.xlsx 作训练集。

4.3 创新点

本文复现主要创新点有两处，其一为应用自动机器学习范式，对原文方法依赖于人类专家经验设置的模型超参数，特别是数据集特定的异常比值 (anomaly ration) r，进行了自动化、并行化、多目标的搜索优化，使得原文方法在训练时更少依赖于人类专家经验，具备更强的适应性。其二为评估了原文方法与复现方法在网络入侵检测时序数据上的效果，为网络入侵检测领域的研究和应用提供有价值的参考。

4.4 基于自动机器学习的超参数搜索优化

在 Anomaly Transformer 模型中，手工设置 anomaly ratio r 依赖于人类实践经验，经过复现评估发现，其设置直接深刻影响到模型在实际应用中的性能。虽然 anomaly ratio 在用于测试时可作为灵敏度调节参数来适应不同情况，但在训练时却必须进行最优搜索设置以实现更好的模型性能。为了解决训练时 anomaly ratio 依赖于人类实践经验的问题，本文采用了自动机器学习的方法，具体而言，使用了 Optuna 这一强大的自动优化框架。

Anomaly ratio 的选择对于模型的性能至关重要，过低的值可能导致漏报率升高，而过高的值则可能使得模型对正常样本产生误报。传统的手动调整方法需要耗费大量的时间和精力，并且可能无法找到全局最优解。因此，引入自动机器学习进行优化是提高效率和性能的重要手段。

Optuna 是一个开源的自动超参数优化框架，其核心思想是通过搜索超参数空间，找到能最小化或最大化定义的目标函数的参数配置。在这个场景下，本文将 anomaly ratio 视为超参数，而目标函数则是模型在验证集中的异常分类评价指标，分别是准确度、精准度、召回率和 F1-Score，为了更全面平衡模型性能，还可以增加训练时间作为一个目标进行最小化。

使用 Optuna 的优势在于其高度灵活的 SOTA 搜索算法和并行化支持，能够更有效地探索超参数空间。通过定义合适的搜索空间和目标函数，Optuna 将在多个试验中进行自适应搜索，最终找到一个最优的超参数设置。这使得模型能够在不同数据分布和业务场景下都取得良好的性能。

整个过程是自动化的，无需手动介入，且 Optuna 实现了可视化控制台用于监视多进程搜索中，各个参数对于目标函数的贡献重要性以及其他超参数与目标函数值的统计参数，从而大大减轻了调整过程的负担。随着数据和业务需求的变化，可以定期运行 Optuna 以重新优化 anomaly ratio，确保模型一直保持在最优状态。

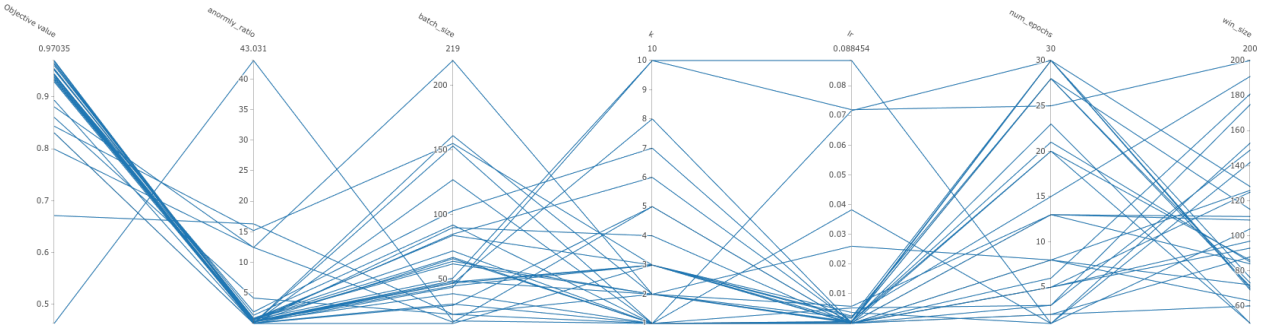
总体而言，通过引入 Optuna 这样的自动机器学习工具，本文希望更有效地解决 Anomaly Transformer 模型中 anomaly ratio 设置的问题，从而提高模型的性能和适应性。

5 实验结果分析

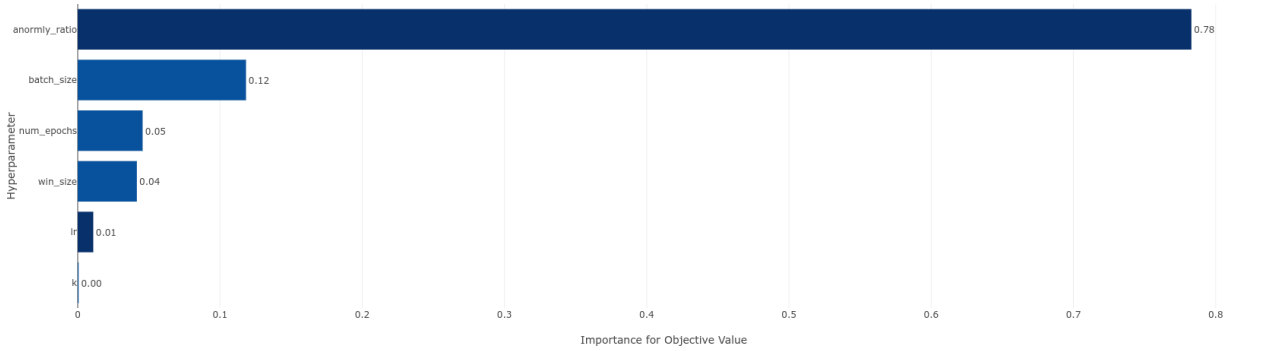
本文安排了三个实验，其一是原论文中在 SMD、MSL、SMAP、SWaT 和 PSM 五个数据集上使用原文 Anomaly Transformer 进行训练和评估所得到的精准度、召回率和 F1-Score 三个评估指标，并绘制分类 ROC 曲线。计算曲线下方面积 AUC。这些评估结果将与原文对比，验证原文实验结果的可靠性。其二是使用 Optuna 框架对 Anomaly Transformer 在 SWaT 数据集上的训练评估超参数进行搜索优化，将优化后的评估指标与原论文的未优化的参数下得到的评估指标进行对比，验证 Optuna 优化效果。其三是使用 Optuna 优化超参数的 Anomaly Transformer 在网络入侵流量数据集上进行异常检测，评估该方法在网络入侵异常检测上的效果。

第一个实验为完全的复现实验，得到图5所示的复现与原文 ROC 曲线与 AUC 值对比图以及表1所示的复现与原文精准度、召回率和 F1-Score 在五个数据集上的效果，对比显示复现结果与原文相比基本一致，但在 SWaT 数据集上的三个指标相对都比原文要低。分析认为，这主要是因为原文作者没有提供对 SWaT 数据集的评估代码所导致的。因此本文安排的第二个实验就在 SWaT 数据集上进行超参数优化。

第二个实验针对 SWaT 数据集进行了模型的超参数，包括 anomaly ratio、学习率、滑动窗口尺寸、batch 尺寸与 k 值进行了共计 30 轮次的搜索优化，得到如图 3b所示的超参数重要性统计柱状图以及图 3a所示的超参数与目标值平行坐标图。其中图 3b所示的柱状图中，可以明显发现 anomaly ratio 的重要性指标为 0.78，远高于其他超参数，表明了 anomaly ratio 对于模型性能的重要影响。图 3a中的目标值是准确率 (权值 0.1)、精准度 (权值 0.3)、召回率 (权值 0.3) 和 F1-Score(权值 0.3) 四个指标的加权求和结果，可以明显发现目标值高的曲线在 anomaly ratio 底部的线条更为密集，这表示当 anomaly ratio 在较小时的目标值更高。最终经过优化搜索，Optuna 找到了最优参数组合，使得 SWaT 数据集的评估实验中精准率达到了 94.12%，召回率达到了 99.07%，F1-Score 达到了 96.53%，比表 1中未优化的原文及复现效果都要好，验证了本文的超参数优化搜索方法的有效性。



(a) 超参数与加权求和目标值的平行坐标图



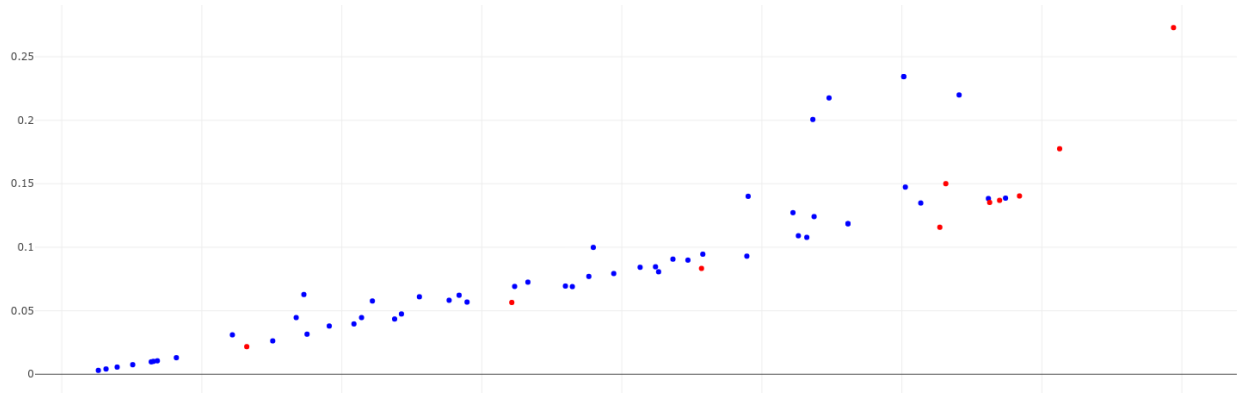
(b) 超参数重要性统计柱状图

图 3. Optuna 优化 Anomaly Transformer 方法在 SWaT 数据集训练与评估统计

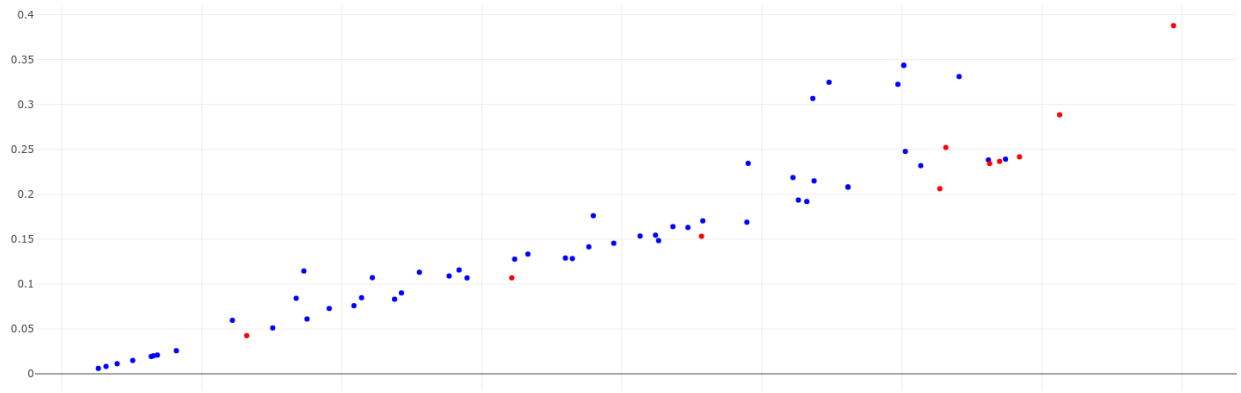
第三个实验是将超参数优化 Anomaly Transformer 扩展应用到网络入侵检测数据集 NSLKDD 上。这个实验中的目标值使用了多目标优化方式，得到了以准确率与 F1-Score、准确率与召回率的帕累托前沿散点图4。其中最右上角的点为同一次实验所得结果，以此为最优参数，得到在 NSLKDD 数据集上的评估结果：精准率为 66.92%，召回率为 27.29%，F1-Score 达到了 38.77%，评估结果并不乐观，因此认为使用超参数优化的 Anomaly Transformer 方法在 NSLKDD 网络入侵检测数据集上并不能取得很好的效果。

表 1. 原文 (O) 与复现 (M) 在五个数据集上的评估实验指标对比。SWaT 数据集 [26] 由于原文未提及具体处理方式也未在开源代码提供具体处理方法与处理结果，因此该数据在复现时未做除标准化外的其他数据处理。

数据集	SMD		MSL		SMAP		SWaT		PSM	
原文/复现	O	M	O	M	O	M	O	M	O	M
精准度	89.40	89.27	92.09	91.86	94.13	93.60	91.55	88.41	96.91	96.97
召回率	95.45	93.29	95.15	95.45	99.40	99.43	96.73	93.71	98.90	98.83
F1-Score	92.33	91.24	93.59	93.62	96.69	96.42	94.07	90.99	97.89	97.89

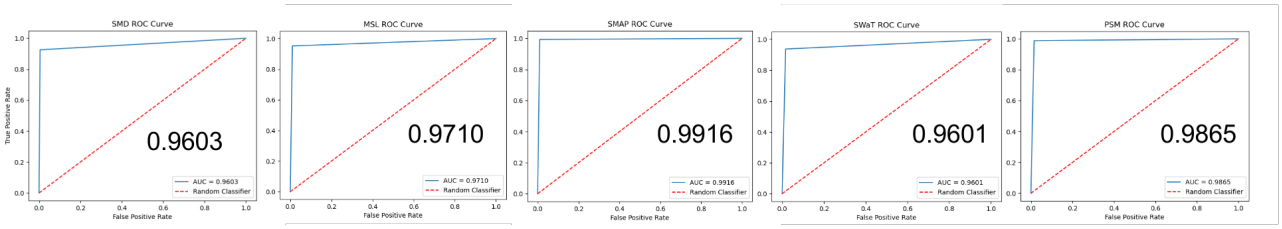


(a) 准确率与召回率的帕累托前沿散点

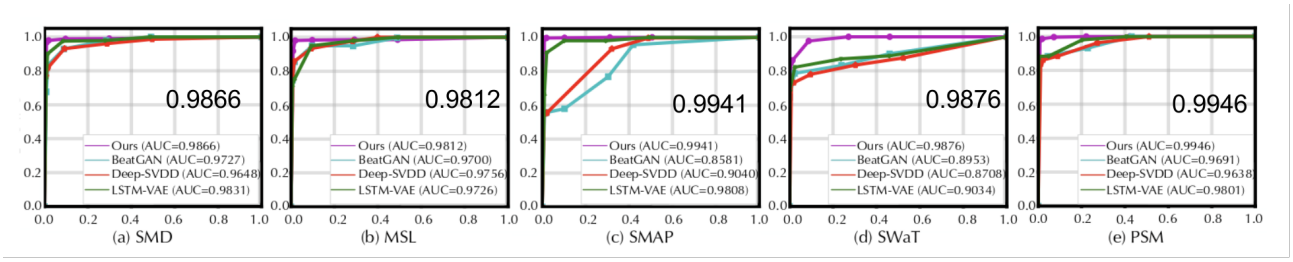


(b) 准确率与 F1-Score 的帕累托前沿散点

图 4. Optuna 优化 Anomaly Transformer 方法在 NSLKDD 数据集上以准确率与 F1-Score、准确率与召回率的帕累托前沿散点图



(a) 本报告复现 ROC 曲线与 AUC 值



(b) 原文 ROC 曲线与 AUC 值

图 5. 原文与复现算法在五个数据集上的异常分类的 ROC 曲线与对应 AUC 值。图 5a与图 5b从左至右均分别对应 SMD、MSL、SMAP、SWAT、PSM 数据集。曲线图中以较大的以黑色字体标记的浮点数为对应曲线的 AUC 值。图 5b中紫色曲线为原文的 ROC 曲线，其余颜色曲线为原文中的对比方法的 ROC 曲线。

6 总结与展望

本文关注了一种基于 Transformer 的时序异常检测方法。该方法的关键创新在于提出了异常点关联集中在相邻时间点的核心观察，并设计了异常注意力机制同时建模时间点的局部和全局关联性。文章定义了关联差异概念，通过两种关联分布的差异来判别正常和异常时间点。为增强关联差异的区分能力，提出了极化训练策略。在多个开源数据集上验证了方法的有效性。为了复现该论文并进行验证，本文了三项主要实验来评价所提出方法的效果:

首先在多个开源数据集上复现了原文方法，结果显示精准度、召回率和 F1 得分与原文基本一致，验证了方法的有效性。其次利用 Optuna 框架对模型超参数进行自动搜索优化，特别是 anomaly ratio 的优化，结果表明明显提高了模型在 SWaT 数据集上的性能，说明超参数优化的重要性。最后在 NSLKDD 网络入侵检测数据集上评估方法效果，但结果并不佳，说明方法在这个场景下效果有限。综上所述，本文通过复现验证了所提出 Anomaly Transformer 在时序异常检测任务上的进步，同时也表明了超参数优化的必要性。但是方法本身在处理不同场景的数据时仍有进一步改进的空间，特别是在网络入侵检测领域。未来工作可考虑引入连续过程建模，处理多变量时间序列，以进一步提升模型的泛化性能。

参考文献

[1] Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy, June 2022.

[2] Donghwoon Kwon, Kathiravan Natarajan, Sang C Suh, Hyunjoo Kim, and Jinoh Kim. An empirical study on network anomaly detection using convolutional neural networks. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, pages 1595–1598. IEEE, 2018.

[3] Mohammad Kazim Hooshmand and Doreswamy Hosahalli. Network anomaly detection using deep learning techniques. *CAAI Transactions on Intelligence Technology*, 7(2):228–243, 2022.

[4] Wenjie Hu, Yihua Liao, and V Rao Vemuri. Robust anomaly detection using support vector machines. In *Proceedings of the international conference on machine learning*, pages 282–289. Citeseer University Park, PA, USA, 2003.

[5] Mohammad Kazim Hooshmand et al. Using ensemble learning approach to identify rare cyber-attacks in network traffic data. In *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 141–146. IEEE, 2020.

[6] Syeda Manjia Tahsien, Hadis Karimipour, and Petros Spachos. Machine learning based solutions for security of internet of things (iot): A survey. *Journal of Network and Computer Applications*, 161:102630, 2020.

[7] Rifkie Primartha and Bayu Adhi Tama. Anomaly detection using random forest: A performance revisited. In *2017 International conference on data and software engineering (ICoDSE)*, pages 1–6. IEEE, 2017.

[8] Sunhee Baek, Donghwoon Kwon, Jinoh Kim, Sang C Suh, Hyunjoo Kim, and Ikkyun Kim. Unsupervised labeling for supervised anomaly detection in enterprise and cloud networks. In *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, pages 205–210. IEEE, 2017.

[9] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.

[10] Jian Tang, Zhixiang Chen, Ada Wai-Chee Fu, and David W Cheung. Enhancing effectiveness of outlier detections for low density patterns. In *Advances in Knowledge Discovery and Data Mining: 6th Pacific-Asia Conference, PAKDD 2002 Taipei, Taiwan, May 6–8, 2002 Proceedings 6*, pages 535–548. Springer, 2002.

[11] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations*, 2018.

[12] Takehisa Yairi, Naoya Takeishi, Tetsuo Oda, Yuta Nakajima, Naoki Nishimura, and Noboru Takata. A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction. *IEEE Transactions on Aerospace and Electronic Systems*, 53(3):1384–1401, 2017.

[13] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54:45–66, 2004.

[14] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018.

[15] Lifeng Shen, Zhuocong Li, and James Kwok. Timeseries anomaly detection using temporal hierarchical one-class network. *Advances in Neural Information Processing Systems*, 33:13016–13026, 2020.

[16] Youjin Shin, Sangyup Lee, Shahroz Tariq, Myeong Shin Lee, Okchul Jung, Daewon Chung, and Simon S Woo. Itad: integrative tensor-based anomaly detection system for reducing false positives of satellite systems. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 2733–2740, 2020.

[17] Daehyung Park, Yuuna Hoshi, and Charles C Kemp. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. *IEEE Robotics and Automation Letters*, 3(3):1544–1551, 2018.

[18] D Li, D Chen, B Jin, L Shi, J Goh, and SK Ng. Madgan: Multivariate anomaly detection for time series data with generative adversarial networks: 703–716, 2019.

[19] Zhihan Li, Youjian Zhao, Jiaqi Han, Ya Su, Rui Jiao, Xidao Wen, and Dan Pei. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 3220–3230, 2021.

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[21] Oliver D Anderson. Time-series. 2nd edn., 1976.

[22] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 387–395, 2018.

[23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[24] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631, 2019.

[25] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A Ghorbani. A detailed analysis of the kdd cup 99 data set. In *2009 IEEE symposium on computational intelligence for security and defense applications*, pages 1–6. Ieee, 2009.

[26] Jonathan Goh, Sridhar Adepu, Khurum Nazir Junejo, and Aditya Mathur. A dataset to support research in the design of secure water treatment systems. In *Critical Information Infrastructures Security: 11th International Conference, CRITIS 2016, Paris, France, October 10–12, 2016, Revised Selected Papers 11*, pages 88–99. Springer, 2017.