Journal of Image and Signal Processing 图像与信号处理, 2024, 13(1), 21-32
Published Online January 2024 in Hans. https://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2024.131003

Hans汉斯

CAE_AD：基于卷积自编码器的无监督时间序列异常检测方法

韩广阳[1]，牛少彰[1,2]，王茂森[2]，史成洁[3]，安洪旭[2]

[1]北京邮电大学计算机学院(国家示范性软件学院)，北京
[2]东南数字经济发展研究院，浙江 衢州
[3]中国科学院信息工程研究所，北京

收稿日期：2023年12月5日；录用日期：2024年1月10日；发布日期：2024年1月17日

摘　要

时间序列数据的有效异常检测对现代工业应用非常重要。然而，由于缺乏异常标签、数据高波动性、训练不稳定，导致建立一个能够准确地进行异常检测的系统是一个具有挑战性的问题。尽管异常检测的深度学习方法最近有所发展，但其中只有少数能够应对所有这些挑战。本文提出了CAE_AD，这是一种基于卷积自编码器(CAE)的无监督异常检测模型。为了尽量地放大异常，避免错过异常，笔者引入了两阶段的对抗训练。同时，为了提高训练稳定性，笔者引入了第一阶段的重建误差以作为第二阶段卷积自编码器的输入。笔者将CAE_AD与先进的时间序列异常检测方法在多个数据集上进行了比较。实验结果表明，本文提出的模型性能优于这些对比方法。在SMAP数据集上，相比于其他模型，CAE_AD模型的f1领先了4%，Precision领先了8%。

关键词

时间序列，异常检测，对抗训练，卷积自编码器

CAE_AD: Unsupervised Time Series Anomaly Detection Method Based on Convolutional Autoencoder

Guangyang Han[1], Shaozhang Niu[1,2], Maosen Wang[2], Chengjie Shi[3], Hongxu An[2]

[1]School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing
[2]Southeast Digital Economy Development Institute, Quzhou Zhejiang
[3]Institute of Information Engineering, Chinese Academy of Sciences, Beijing

文章引用: 韩广阳, 牛少彰, 王茂森, 史成洁, 安洪旭. CAE_AD: 基于卷积自编码器的无监督时间序列异常检测方法[J]. 图像与信号处理, 2024, 13(1): 21-32. DOI: 10.12677/jisp.2024.131003

Received: Dec. 5th, 2023; accepted: Jan. 10th, 2024; published: Jan. 17th, 2024

Abstract

Effective anomaly detection of time series data is crucial for modern industrial applications. However, due to the lack of anomaly labels, high data volatility, and unstable training, establishing a system that can accurately detect anomalies is a challenging problem. Although deep learning methods for anomaly detection have recently developed, only a few of them can address all of these challenges. This article proposes CAE_AD, which is an unsupervised anomaly detection model based on convolutional autoencoder (CAE). In order to maximize the amplification of anomalies and avoid missing them, the author introduced two-stage adversarial training. Meanwhile, in order to improve training stability, the author introduced the reconstruction error from the first stage as the input for the convolutional autoencoder in the second stage. The author compared CAE_AD with advanced time series anomaly detection methods on multiple data sets. The experimental results show that the model proposed in this article performs better than these comparison methods. On the SMAP dataset, compared to other models, the CAE_AD model has a 4% lead in f1 and an 8% lead in Precision.

Keywords

Time Series, Abnormal Detection, Adversarial Training, Convolutional Autoencoder

1. 引言

近几十年来，异常检测一直是机器学习领域的核心研究之一，具有广泛的应用。异常检测似乎是一个简单的两类分类，即可以利于对正常或异常数据进行分类。然而，它也面临着以下挑战。

首先，训练数据正负样本高度不平衡。因为与正常数据相比，数据集中的异常数据通常非常罕见。其次，用户没有简单的方法来手动标记每个训练数据，特别是异常标签。由于上述挑战，与半监督和监督学习方法相比，使用无监督学习方法进行异常检测的趋势越来越大，因为无监督方法可以处理不平衡和未标记的数据。

另外，工业场景中，物联网平台中的传感器和设备数量不断增加，同时数据具有高的波动性，导致异常检测模型的训练稳定性越来越差。

现有解决方案。上述挑战导致开发了无数用于自动异常检测的无监督学习解决方案。研究人员开发了基于重建的方法，其主要目的是封装时间趋势，并以无监督的方式重建时间序列数据，然后使用重建数据与真实数据的差值作为异常分数。传统方法，如基于密度的时间序列异常检测方法 LOF [1]、COF [2]，基于聚类的时间序列异常检测方法 DBSCAN [3]、CBLOF [4]，基于距离的时间序列异常检测方法 KNN [5]、K-means [6]，基于树的时间序列异常检测方法 iForest [7]、决策树[8]。基于循环神经网络 RNN [9]的方法可以捕捉时间戳之间的依赖关系，但是训练 RNN 需要花费大量的时间，效率低。MAD_GAN [10]使用生成对抗网络(GAN)进行异常检测，但是训练 GAN 会存在不收敛的问题。LSTM-NDT [11]方法使用基于长短期记忆网络(LSTM)来预测数据，并使用输入时间序列和非参数动态

阈值方法来检测预测误差中的异常，然而 LSTM 模型在面临高波动性的数据时会出现不稳定性和梯度消失现象。

针对生成对抗网络 GAN 训练时不收敛、不稳定的问题，笔者设计了基于卷积自编码器 CAE 的无监督时间序列异常检测方法 CAE_AD，该方法受到生成对抗网络(GAN)的启发，基于卷积自编码器架构。CAE_AD 基于编码器－解码器架构的对抗性训练，使用重建误差使其能够学习如何放大微弱异常。本文的主要贡献如下：

(1) 提出了基于卷积自编码器(CAE)的无监督时间序列异常检测模型，以解决实际应用中异常标签少、正负样本不均衡的问题。

(2) 如果重建误差太小，即重建值相对接近正常数据，卷积自编码器(CAE)往往会错过异常。将卷积自编码器(CAE)和对抗训练相结合，可以放大重建偏差，避免错过微弱的异常。

(3) 将第一阶段的重建误差作为第二阶段的输入，可以提高模型的训练稳定性，以解决数据高波动性导致训练不稳定的问题。

2. 相关工作

时间序列异常检测问题在学术界和工业界研究已久，关于时间序列异常检测的方法有很多，包括基于密度、聚类、距离、树等传统的异常检测方法。近年来，基于深度学习的时间序列异常检测方法日益成为研究的重点，LSTM_NDT [11]、OmniAnomaly [12]、DAGMM [13]、MAD_GAN [10]、USAD [14]是近年来性能比较好的深度学习异常检测方法，本文以这五个模型作为实验对比模型。LSTM_NDT 模型[11]使用长短期记忆网络(LSTM)模型来预测输入数据下一个时间戳的数据，并且提出了一种非参数异常阈值方法NDT，然而LSTM模型在面临高波动性的数据时会出现不稳定性和梯度消失现象。OmniAnomaly [12]是一种用于多变量时间序列异常检测的随机递归神经网络，其核心思想是利用随机变量连接和平面归一化流等关键技术，通过学习多变量时间序列的鲁棒表示，捕捉多变量时间系列的正态模式，通过表示重构输入数据，并利用重构概率确定异常。DAGMM [13]是一种用于无监督异常检测的深度自编码高斯混合模型，为每个输入数据点生成低维表示和重建误差，并将其进一步输入到高斯混合模型。DAGMM 以端到端的方式同时联合优化深度自动编码器和混合模型的参数，利用单独的估计网络来促进混合模型的参数学习，然而 DAGMM 很慢，并且不能很好地利用模态间的相关性。MAD_GAN [10]是一种基于生成对抗网络(GAN)的无监督多变量异常检测方法，使用长短期记忆递归神经网络(LSTM-RNN)作为 GAN 框架中的基本模型(即生成器和鉴别器)来捕获时间序列分布的时间相关性，使用一种称为 DR 评分的新异常评分来通过判别和重建来检测异常，但是训练 GAN 会存在不收敛的问题。USAD [14]使用两个自编码器通过对抗性训练来实现异常检测，其自动编码器架构使其能够以无监督的方式进行学习，对抗性训练及其体系结构的使用使其能够在提供快速训练的同时隔离异常，然而在训练高波动性的数据时存在训练不稳定的问题。

针对以上方法存在容易错过微弱异常、训练不收敛和训练不稳定的问题，笔者设计了基于卷积自编码器 CAE 的无监督时间序列异常检测方法 CAE_AD。针对存在容易错过微弱异常的问题，笔者将卷积自编码器和两阶段的对抗训练相结合，放大第一阶段的重建误差，避免错过微弱的异常；针对 MAD_GAN 和 USAD 训练过程中的不收敛和不稳定问题，笔者将第一阶段的重建误差作为第二阶段的输入，提高模型的训练稳定性，以解决数据高波动性导致训练不稳定的问题。最后，在公共数据集上的实验对比分析表明了所提模型的有效性。

3. 基于卷积自编码器的无监督异常检测方法

本节中，将从时间序列异常检测的问题描述、数据处理、两阶段的对抗训练、异常分数的计算方法

四个方面来介绍基于卷积自编码器的无监督异常检测方法。

3.1. 时间序列异常检测的问题描述

在时间序列异常检测中，数据通常是带有时间戳的长度为 T 的一系列数据点

$$T=\{x_1,...,x_T\} \tag{1}$$

其中，每一个数据点 x_t 在特定的时间戳 t 被物理传感器采集，$x_t \in R^m$。如果 $m>1$，此时为多变量时间序列数据；如果 $m=1$，是一种特殊情况，此时是单变量时间序列数据。在本文中，我们考虑使用无监督的方法做时间序列异常检测。训练时，给定训练输入时间序列 T，假设 T 只包含正常数据，无异常数据。测试时，对于长度为 $\hat{L}$ 的测试数据 $\hat{T}$，我们重建 $\hat{T}$ 得到模型的输出 O，如果测试数据 $\hat{T}$ 与模型输出 O 的差值大于阈值 D，则检测为异常值。通过上述方法，我们得到检测结果 $Y=\{y_1,...,y_{\hat{L}}\}$，使用 $y_t \in \{0,1\}$ 来表示测试集的时间戳 t 数据是否为异常值。y_t 为 1 表示为异常数据点，此时测试数据 $\hat{T}_t$ 与模型输出 O_t 的差值大于等于阈值 D；y_t 为 0 表示为正常数据点，此时测试数据 $\hat{T}_t$ 与模型输出 O_t 的差值小于阈值 D。

3.2. 数据处理

为了提高模型的收敛速度，防止模型梯度爆炸，使用线性归一化方法对数据进行了归一化操作。线性归一化方法的公式如下

$$x_t \leftarrow \frac{x_t-\min(T)}{\max(T)-\min(T)} \tag{2}$$

其中，$\min(T)$ 是训练数据中的最小向量值，$\max(T)$ 是训练数据中的最大向量值。经过过一化后，全部数据处于[0, 1]范围内。为了捕捉时间戳之间的依赖关系，提高模型的准确性，将某个时间戳的数据扩大到包含这个时间戳的一段区间。对于时间戳 t，定义长度为 K 的时间窗口 W_t：

$$W_t=\{x_{t-K+1},...,x_{t-1},x_t\} \tag{3}$$

如果 $t<K$，使用长度为 $K-t$ 的向量 $\{x_t,...,x_t\}$ 追加填充在 W_t 的尾部，以维持 W_t 的长度为 K。定义滑动窗口序列 W 作为模型训练的输入数据，W 定义如下：

$$W=\{W_1,...W_T\} \tag{4}$$

在测试时，使用 $\widehat{W}$ (长度为 $\hat{L}$)作为测试的输入数据，通过本文提出的模型，得到重建值 O。在 t 时间，使用输入数据 $\widehat{W}_t$ 和重建值 O_t 来计算异常分数 S_t，如果异常分数 S_t 大于阈值 D，则检测结果 $y_t=1$ (检测为异常)，否则检测结果 $y_t=0$ (检测为正常)。

3.3. 两阶段的对抗训练

本文中的 CAE_AD 模型基于两阶段对抗训练的卷积自编码器(CAE)架构，如图 1 所示。CAE_AD 由三个部分组成：一个编码器网络 Encoder 和两个解码器网络 Decoder1 和 Decoder2。如图 2 所示，有两个卷积自编码器 CAE1 和 CAE2，共享编码器网络 Encoder。CAE_AD 分为两个阶段进行训练。首先，训练两个 CAE 来完成对输入窗口 W 的重建。然后，CAE1 和 CAE2 进行对抗式的训练，其中 CAE1 尽可能地生成与真实数据 W 接近的数据来试图欺骗 CAE2，而 CAE2 的目标是学习数据是真实的(直接来自 W)还是重构生成的(来自 CAE1)。

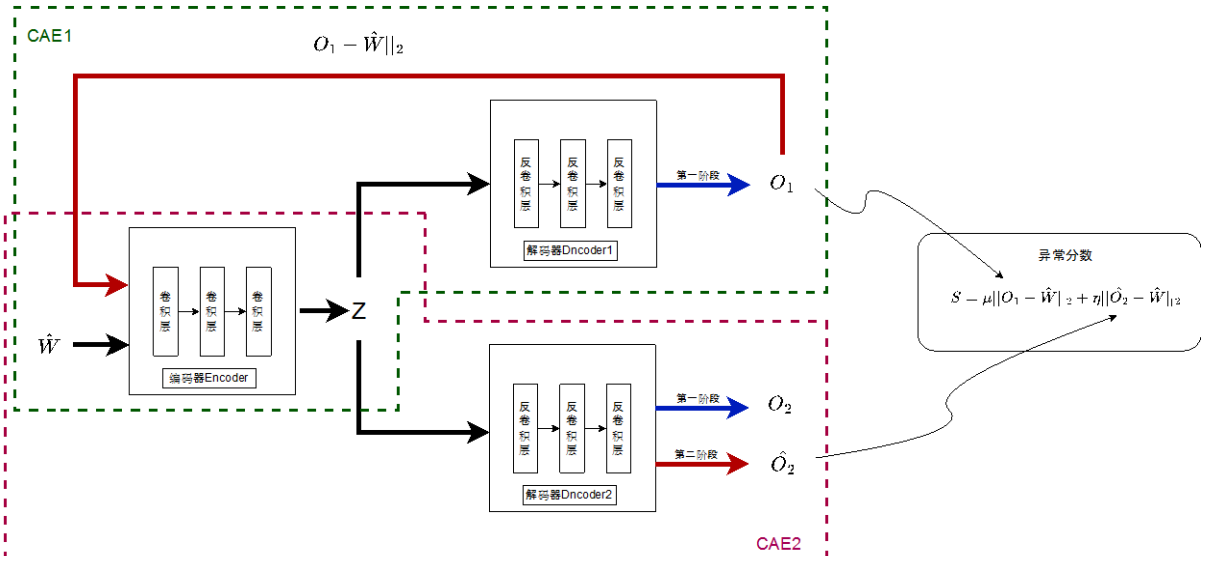

Figure 1. CAE_AD model diagram
图 1. CAE_AD 模型图

第一阶段：卷积自编码器训练。在第一阶段，目的是训练 CAE1 和 CAE2 学习重建输入窗口 W。输入窗口 W 被编码器 Encoder 压缩到潜在变量 Z，然后由解码器 Decoder1 和 Decoder2 将潜在变量 Z 还原到和输入窗口 W 一样的维度，从而得到重建值 O_1 和 O_2。这个过程可由如下公式表示：

$$O_1 = D_1\left(E\left(W\right)\right) \tag{5}$$

$$O_2 = D_2\left(E\left(W\right)\right) \tag{6}$$

在第一阶段，使用 L2 范数定义两个解码器的重建损失，如下：

$$L_1 = \left\|O_1 - W\right\|_2 \tag{7}$$

$$L_2 = \left\|O_2 - W\right\|_2 \tag{8}$$

第二阶段：两阶段对抗性训练。目的是训练 CAE2 区分真实数据和来自 CAE1 的重建数据，训练 CAE1 欺骗 CAE2。使用 $\left\|O_1 - W\right\|_2$ 来作为第一阶段的重建误差。在第二阶段，使用第一阶段的重建误差作为第二阶段的输入，重建误差被编码器 Encoder 压缩到潜在空间 Z，然后被 Decoder2 重建得到 $\widehat{O_2}$。第二阶段的重建值 $\widehat{O_2}$ 可由如下公式表示：

$$\widehat{O_2} = D_2\left(E\left(\left\|O_1 - W\right\|_2\right)\right) \tag{9}$$

在对抗训练中，CAE1 目的是最小化第二阶段重建输出 $\widehat{O_2}$ 和输入窗口 W 之间的差异，通过完美地重建输入(即 O_1 约等于 W，相差很小)来创建一个重建误差 $\left\|O_1 - W\right\|_2$，从而欺骗 CAE2。而 CAE2 的目的是最大限度地扩大第二阶段重建输出 $\widehat{O_2}$ 和输入窗口 W 之间的差异。训练 CAE1 是否成功欺骗了 CAE2，训练 CAE2 是否可以将 CAE1 的重建数据 O_1 与真实数据 W 区分开来。训练目标是

$$\min_{CAE1}\max_{CAE2}\left\|\widehat{O_2} - W\right\|_2 \tag{10}$$

CAE1 的目标是使 $\left\|\widehat{O_2} - W\right\|_2$ 最小化，而 CAE2 的目标是使 $\left\|\widehat{O_2} - W\right\|_2$ 最大化。通过使用损失来实现：

$$L_1 = +\left\|\widehat{O_2} - W\right\|_2 \tag{11}$$

$$L_2 = -\left\|\widehat{O_2} - W\right\|_2 \tag{12}$$

现在有了第一阶段的重建损失和第二阶段的对抗性损失，可以得到 CAE1 和 CAE2 的两阶段累计损失函数。具体的损失函数如下：

$$L_1 = \frac{1}{n}\left\|O_1 - W\right\|_2 + \left(1 - \frac{1}{n}\right)\left\|\widehat{O_2} - W\right\|_2 \tag{13}$$

$$L_2 = \frac{1}{n}\left\|O_2 - W\right\|_2 - \left(1 - \frac{1}{n}\right)\left\|\widehat{O_2} - W\right\|_2 \tag{14}$$

n 表示当前训练的轮次，从 1 开始递增。在训练的初始阶段，n 比较小，$1/n$ 比较大，$(1-1/n)$比较小。即刚开始训练时，让重建损失的权重很高，对抗性损失的权重很低，目的是保证第一阶段重建效果不佳时(第一阶段的重建值 O_1 和输入窗口 W 的值相差较大)的训练稳定性。在第一阶段重建不佳的情况下，那么作为第二阶段输入的重建误差 $\left\|O_1 - W\right\|_2$ 将不可靠。因此，在训练的初始阶段，对抗性损失被赋予较低的权重，以防止出现训练不稳定的问题。随着训练的轮数 n 不断增加，第一阶段的重建值 O_1 越来越接近输入窗口 W，使得重建误差 $\left\|O_1 - W\right\|_2$ 越来越准确，对抗性损失的权重也在增加。

3.4. 异常分数的计算方法

在异常检测的测试阶段，对于测试集输入数据 $\widehat{W}$，异常分数被定义为

$$S = \mu\left\|O_1 - \widehat{W}\right\|_2 + \eta\left\|\widehat{O_2} - \widehat{W}\right\|_2 \tag{15}$$

其中 $\mu + \eta = 1$，可以动态调整异常分数中的第一阶段和第二阶段的权重，使得模型对两个阶段的灵敏度不同，以适应不同的应用场景。

在测试阶段，如果异常分数大于阈值，则标记当前时间戳 t 为异常值。使用 POT [15]方法来自动和动态地选择阈值，POT 使用“极值理论”将数据分布与广义帕累托分布拟合，并确定适当的风险值，以动态确定阈值。

数据的维度是 m，每个维度的异常诊断标签 y_i 和检测结果 y 定义为

$$y_i = 1\left(S_i \geq POT\left(S_i\right)\right) \tag{16}$$

$$y = \bigvee_i y_i \tag{17}$$

只要 m 维中任意一个维度为异常的，那么就将当前时间戳标记为异常。

4. 实验结果和分析

本节中，在 4.1 中介绍了实验用到的三个公开数据集，在 4.2 中介绍了实验用到的评价指标，在 4.3 中分析了 CAE_AD 模型中四个参数对实验结果的影响，在 4.4 中 CAE_AD 使用 4.3 中得到的性能最优的四个参数，与 MAD_GAN、LSTM_NDT、DAGMM、OmniAnomaly、USAD 模型进行了性能对比，在 4.5 中对 CAE_AD 模型进行了消融实验。

4.1. 公开数据集

在实验中，使用了 3 个公开的数据集：SMAP、SWaT、MSL。在表 1 中总结了 3 个数据集的基本情况，包括训练集、测试集的大小，维度，异常率，括号中的值是数据集中的实体数。比如，MSL 数据集有 27 个实体，每个实体有 55 个维度。下面详细介绍上述 3 个数据集的情况。

Table 1. Dataset situation
表 1. 数据集情况

数据集	训练集大小	测试集大小	维度	异常率(%)
SMAP	135183	427617	25 (55)	13.13
SWaT	496800	449919	51 (1)	11.98
MSL	58317	73729	55 (27)	10.72

土壤水分监测卫星时序数据集(SMAP)和火星科学实验室探测器“好奇号”时序数据集(MSL)是美国国家航天航空局 NASA 的卫星公开时序数据集。SMAP 数据集有 55 个实体，每个实体包含 25 个维度，整体异常率为 13.13%；MSL 数据集有 27 个实体，每个实体包含 55 个维度，整体异常率为 10.72%。

安全水处理(SWaT)数据集源自于新加坡公用事业委员会协调的运行水处理试验台。数据收集了持续四天的 51 个传感器记录，频率为 1 秒。总共进行了 36 次攻击，导致 11.98%的时间出现了异常状况。因此，SWaT 数据集有一个实体，这个实体包含 51 个维度，整体异常率有 11.98%。

4.2. 评价指标

在介绍用于评估时间序列异常检测性能的指标之前，首先介绍 4 个符号：TP (真阳性)、FP (假阳性)、FN (假阴性)、TN (真阴性)。TP 是所有被正确预测为正的样例数；FP 是所有被错误预测为正的样例数；TN 是所有被正确预测为负的样例数；FN 是所有被错误预测为负的样例数。

用于评估时间序列异常检测性能的指标有精确率 Precision (P)，召回率 Recall (R)，和 F1score (F1)。

(1) 精确率(查准率)：所有被正确预测为正的样例占所有被预测为正的样例的比重。

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{18}$$

(2) 召回率(查全率)：所有被正确预测为正的样例占所有实际为正的样例的比重。

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{19}$$

(3) $F1$：精确率 P 和召回率 R 的调和均值。简单来说，$F1$ 越大，precision 和 recall 的值越大，即 $F1$ 越大越好。

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{20}$$

4.3. 参数影响

本节中，为使 CAE_AD 模型尽可能地取得最优的性能，将分析 CAE_AD 模型中四个重要参数对实验结果的影响。接下来，将从滑动窗口大小、编码器输出通道数、卷积核大小、异常分数参数四个方面来分别分析其对实验结果的影响。

4.3.1. 滑动窗口大小

滑动窗口的大小 K 在重建时间序列数据，以及在判断异常结果时发挥着重要的作用。同时，窗口大小 K 定义了输入数据 W 的长度，可以直接影响异常检测的速度。图 2 显示了 CAE_AD 模型使用不同的窗口大小在 SMAP 数据集上取得的异常检测性能结果，K 取[10, 25, 50, 75, 100]。如果滑动窗口的大小 K 过小，则不能很好地反映时间序列数据在时间维度的特征；如果滑动窗口的大小 K 过大，则会导致异常检测模型在大时间跨度对时间序列数据建模缺乏敏感性，以及导致模型重建的难度大大增加，同时大大降低了异常检测的速度。因此，对于 SMAP 数据集，一个长度适中的窗口大小更利于 CAE_AD 模型发挥

出最优的异常检测性能。由图 2 可知，当窗口大小 K 等于 25 时，异常检测的性能最优，取得了最好的 F1、AUC、Precision 指标效果。

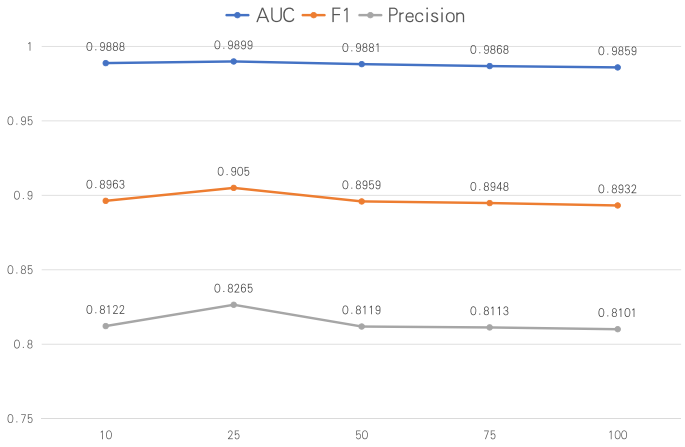

Figure 2. The effect of sliding window size on experimental performance
图 2. 滑动窗口大小对实验性能的影响

4.3.2. 编码器输出通道数

接下来，研究编码器中最后一个卷积层的输出通道数 Z。卷积核的通道数是卷积操作中的一个重要参数，可以影响异常检测模型的性能和计算速度。合理地选择卷积核通道数，可以让卷积操作更有效，提高异常检测模型的表征能力。

图 3 显示了 CAE_AD 模型使用不同的输出通道数 Z 在 SMAP 数据集上取得的异常检测性能结果，Z 取[16, 32, 64, 128, 256]。通过增加卷积核通道数可以扩大卷积核在特征空间的搜索范围，从而能够更好地捕捉时间序列数据的局部特征，以提高异常检测模型的表征能力。如果卷积核通道数 Z 过小，那么卷积输出的特征可能无法包含数据的全部信息，导致模型出现欠拟合的现象；如果卷积核通道数 Z 过大，则模型可能会出现过拟合的现象。因此，编码器中最后一个卷积层输出通道数 Z 的选择需要在欠拟合和过拟合之间寻求一个平衡点，以达到最优的异常检测性能。由图 3 可知，输出通道数 Z 等于 32 时，CAE_AD 模型异常检测的性能最优。

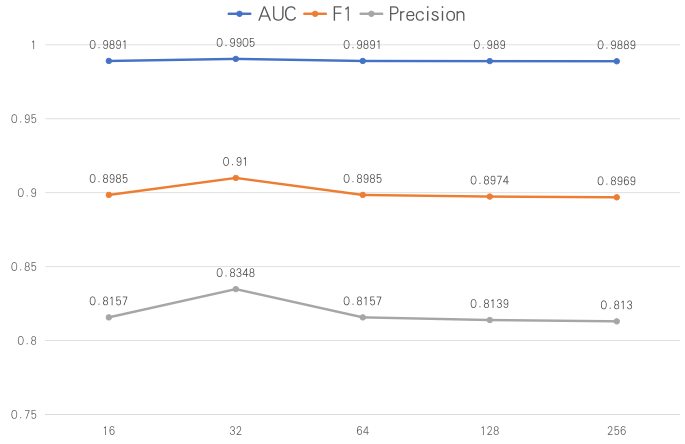

Figure 3. The influence of encoder output channel number on experimental performance
图 3. 编码器输出通道数对实验性能的影响

4.3.3. 卷积核大小

卷积层中的卷积核大小对异常检测模型的检测性能非常重要。图 4 显示了 CAE_AD 模型使用不同的卷积核大小在 SMAP 数据集上取得的异常检测性能结果，卷积核大小取[(2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (7, 7), (8, 8), (9, 9)]。增加卷积核的大小可以让模型有更大的感受野，有利于进行特征提取。如果卷积核过大，会使整个模型的复杂度急剧增加；如果卷积核过小，又会使训练的速度降低。因此，选择大小合适的卷积核对异常检测的性能非常重要。由图 4 可知，卷积核大小等于(3, 3)时，CAE_AD 模型异常检测的性能最优。

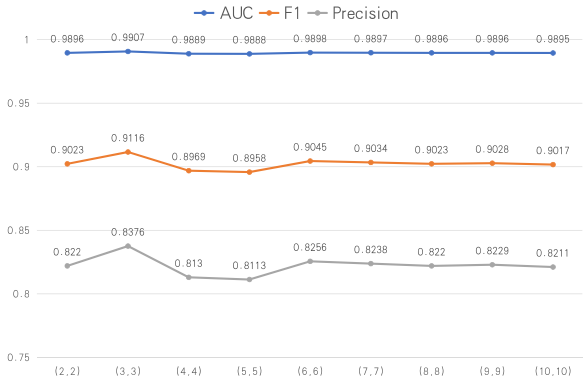

Figure 4. The effect of convolutional kernel size on experimental performance
图 4. 卷积核大小对实验性能的影响

4.3.4. 异常分数的参数

最后，研究了异常分数中的两个参数(μ, η)对异常检测性能的影响。如果 μ 大于 η，则说明模型更注重第一阶段的重建；如果 μ 小于 η，则说明模型更注重第二阶段的重建。可以动态地调整(μ, η)，以改变模型对第一阶段和第二阶段的重视程度，进而使 CAE_AD 模型具有不同的灵敏度，以适应不同生产环境的要求。

图 5 显示了 CAE_AD 模型使用不同的(μ, η)在 SMAP 数据集上取得的异常检测性能结果，(μ, η)取[(0.0, 1.0), (0.1, 0.9), (0.2, 0.8), (0.3, 0.7), (0.4, 0.6), (0.5, 0.5), (0.6, 0.4), (0.7, 0.3), (0.8, 0.2), (0.9, 0.1), (1.0, 0.0)]。使用不同的(μ, η)，可以使 CAE_AD 模型实现不同级别的灵敏度，提高了异常检测的灵活性。由图 5 可知，当(μ, η)等于(0.3, 0.7)时，CAE_AD 模型异常检测的性能最优。

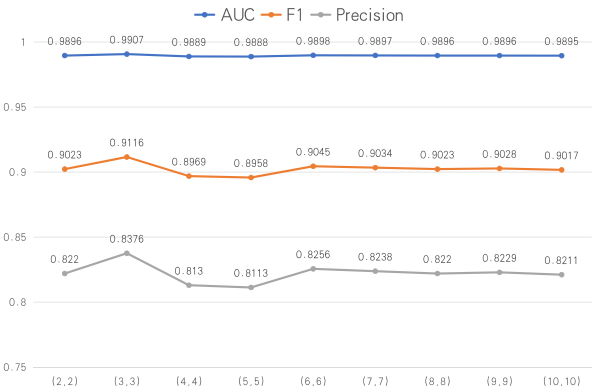

Figure 5. The effect of anomalous score parameters on experimental performance
图 5. 异常分数参数对实验性能的影响

4.4. 对比实验结果

本节中，使用 4.3 中得到的最优参数，滑动窗口大小为 25，编码器输出通道数为 32，卷积核大小为(3, 3)，异常分数参数(μ, η)为(0.3, 0.7)，将 CAE_AD 模型与 MAD_GAN、LSTM_NDT、DAGMM、OmniAnomaly、USAD 等模型在 SMAP、SWaT、MSL 数据集上进行了性能对比。

Table 2. Performance comparison of six models on SMAP, SWaT, and MSL datasets
表 2. 六种模型在 SMAP、SWaT、MSL 数据集上的性能对比

模型	SMAP				SWaT				MSL			
	p	r	AUC	F1	p	r	AUC	F1	p	r	AUC	F1
MAD_GAN	0.8077	0.9999	0.9885	0.8936	0.9866	0.6879	0.8433	0.8106	0.9548	0.9999	0.9763	0.9769
LSTM_NDT	0.7879	0.7326	0.8597	0.7834	**0.9999**	0.0109	0.5054	0.0215	0.7889	0.9999	0.8659	0.8821
DAGMM	0.8069	0.9999	0.9885	0.8931	0.9866	0.6879	0.8433	0.8106	0.9686	0.9999	0.9837	0.984
OmniAnomaly	0.8043	0.9999	0.9882	0.8915	0.9837	0.6698	0.8326	0.8032	0.9677	0.9999	0.9833	0.9836
USAD	0.8095	0.9999	0.9886	0.8947	0.9977	0.677	0.8384	0.8066	0.9686	0.9999	0.9837	0.984
CAE_AD	**0.8884**	**0.9999**	**0.9939**	**0.9409**	0.9739	**0.6957**	**0.8464**	**0.8116**	**0.9752**	**0.9999**	**0.9872**	**0.9874**

表 2 详细说明了在公开数据集上所有时间序列异常检测方法取得的性能结果。在 F1 指标上，CAE_AD 模型的表现在三个公开数据集上优于所有对比的方法，尤其在 SMAP 数据集上，CAE_AD 模型的 F1 是 94.09%，比对比方法中效果最好的 USAD 模型(F1 是 89.47%)提高了 4.6%。在召回率 R 和 AUC 这两个指标上，CAE_AD 在三个公开数据集上同样取得了最好的成绩，优于对比的 5 个异常检测方法。在精确率 P 指标上，在 SMAP 和 MSL 数据集上 CAE_AD 模型优于其他对比方法；在 SWaT 数据集上，LSTM_AD 的精确率 P 最高(99.99%)，CAE_AD 精确率 P 是 97.39%。

4.5. 消融实验

为了研究 CAE_AD 模型中每个部分对于异常检测结果的影响，设计 CAE_AD 模型的三种消融变体，以进一步说明 CAE、重建误差、对抗性训练在异常检测中起到的重要作用。

第一个变体 CAE_AD_1：相对于 CAE_AD 模型，将 CAE 中的卷积模块换成了线性层，以研究 CAE 中卷积模块对异常检测结果的影响。

第二个变体 CAE_AD_2：相对于 CAE_AD 模型，删除了第一阶段的重建误差，而使用第一阶段的输出 O1 作为第二阶段的输入，以研究重建误差模块对异常检测结果的影响。

第三个变体 CAE_AD_3：相对于 CAE_AD 模型，删除了对抗性训练，即使用第一阶段的重建误差进行单一阶段的推理，以研究对抗训练对异常检测结果的影响。

实验结果如表 3 所示，在 SMAP 数据集上三个消融变体和 CAE_AD 的对比效果最为明显。

Table 3. Ablation Research: Performance comparison of CAE_AD and its ablation variants
表 3. 消融实验：CAE_AD 及其消融变体的性能对比

模型	SMAP				SWaT				MSL			
	p	r	AUC	F1	p	r	AUC	F1	p	r	AUC	F1
CAE_AD_1	0.8157	0.9999	0.9891	0.8985	0.9545	0.6848	0.8340	0.7975	0.9653	0.9999	0.9820	0.9823
CAE_AD_2	0.8433	0.9999	0.9910	0.9150	0.9718	0.6957	0.8463	0.8109	0.9735	0.9999	0.9864	0.9866
CAE_AD_3	0.8122	0.9999	0.9888	0.8963	0.9718	0.6957	0.8463	0.8109	0.9735	0.9999	0.9864	0.9866
CAE_AD	**0.8884**	**0.9999**	**0.9939**	**0.9409**	**0.9739**	**0.6957**	**0.8464**	**0.8116**	**0.9752**	**0.9999**	**0.9872**	**0.9874**

以下结论以 SMAP 数据集上的效果为根据：

(1) 就 F1 指标而言，第三个变体 CAE_AD_3 的性能下降最快，大约 4.5%，这表明对抗性训练对于异常检测结果的作用非常关键。如果重建偏差太小，即相对接近正常数据，模型往往会错过异常，对抗性训练可以放大重建偏差，以缓解上述问题。

(2) 当使用第二个变体 CAE_AD_2，即删除重建误差后，F1 指标性能下降了大约 2.5%，这表明重建误差可以显著提高第二阶段的重建性能。将第一阶段的重建残值作为第二阶段的输入，可以提高模型的训练稳定性，以解决数据高波动性导致训练不稳定的问题。

(3) 当第一个变体 CAE_AD_1，即将 CAE 中的卷积模块换成线性模块，F1 指标下降了 4.24%，这表明卷积模块对异常检测的结果非常重要。

总之，CAE_AD 模型中任何模块的消融都会导致较差的结果。

5. 结束语

在本文中，提出了基于卷积自编码器(CAE)的无监督时间序列异常检测方法 CAE_AD。CAE_AD 模型使用卷积自编码器来实现对输入数据的重建，通过无监督异常检测的方式解决实际应用中异常标签少、正负样本不均衡的问题。CAE_AD 模型通过两阶段的对抗训练来放大第一阶段的重建误差，避免错过微小的异常。此外，CAE_AD 模型将第一阶段的重建误差作为第二阶段的输入，从而提高模型的训练稳定性，以解决数据高波动性导致训练不稳定的问题。在实验研究中，CAE_AD 模型在三个公共数据集上表现出优秀的性能，相对于对比方法，F1 分数领先了 4%，Precision 领先了 8%。CAE_AD 模型为在高维数据或者高波动性数据上进行无监督异常检测提供了一个有希望的方向。

参考文献

[1] Huang, T., Zhu, Y., Zhang, Q., *et al*. (2013) Anlof-Based Adaptive Anomaly Detection Scheme for Cloud Computing. 2013 *IEEE* 37*th Annual Computer Software and Applications Conference Workshops*, Japan, 22-26 July 2013, 206-211. https://doi.org/10.1109/COMPSACW.2013.28

[2] Zhang, L., Chen, Y. and Liao, S. (2018) Algorithm Optimization of Anomaly Detection Based on Data Mining. 2018 10*th International Conference on Measuring Technology and Mechatronics Automation* (*ICMTMA*), Changsha, 10-11 February 2018, 402-404. https://doi.org/10.1109/ICMTMA.2018.00104

[3] Çelik, M., Dadaşer-Çelik, F. and Dokuz, A.Ş. (2011) Anomaly Detection in Temperature Data Using DBSCAN Algorithm. 2011 *International Symposium on Innovations in Intelligent Systems and Applications*, Istanbul, 15-18 June 2011, 91-95. https://doi.org/10.1109/INISTA.2011.5946052

[4] He, Z., Xu, X. and Deng, S. (2003) Discovering Cluster-Based Local Outliers. *Pattern Recognition Letters*, **24**, 1641-1650. https://doi.org/10.1016/S0167-8655(03)00003-5

[5] Su, M.Y. (2011) Real-Time Anomaly Detection Systems for Denial-of-Service Attacks by Weighted K-Nearest-Neighbor Classifiers. *Expert Systems with Applications*, **38**, 3492-3498. https://doi.org/10.1016/j.eswa.2010.08.137

[6] Münz, G., Li, S. and Carle, G. (2007) Traffic Anomaly Detection Using K-Means Clustering. *Giitg Workshop Mmbnet*, **7**.

[7] Liu, F.T., Ting, K.M. and Zhou, Z.H. (2012) Isolation-Based Anomaly Detection. *ACM Transactions on Knowledge Discovery from Data* (*TKDD*), **6**, 1-39. https://doi.org/10.1145/2133360.2133363

[8] Muniyandi, A.P., Rajeswari, R. and Rajaram, R. (2012) Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree Algorithm. *Procedia Engineering*, **30**, 174-182. https://doi.org/10.1016/j.proeng.2012.01.849

[9] Nanduri, A. and Sherry, L. (2016) Anomaly Detection in Aircraft Data Using Recurrent Neural Networks (RNN). 2016 *IEEE Integrated Communications Navigation and Surveillance* (*ICNS*), Herndon, VA, 19-21 April 2016, 5C2-1-5C2-8. https://doi.org/10.1109/ICNSURV.2016.7486356

[10] Li, D., Chen, D., Jin, B., *et al*. (2019) MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks. *International Conference on Artificial Neural Networks*, Springer International Publishing, Cham, 703-716. https://doi.org/10.1007/978-3-030-30490-4_56

[11] Hundman, K., Constantinou, V., Laporte, C., *et al*. (2018) Detecting Spacecraft Anomalies Using LSTMS and Nonpa-

rametric Dynamic Thresholding. *Proceedings of the* 24*th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 387-395. https://doi.org/10.1145/3219819.3219845

[12] Su, Y., Zhao, Y., Niu, C., *et al*. (2019) Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. *Proceedings of the* 25*th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, 4-8 August 2019, 2828-2837. https://doi.org/10.1145/3292500.3330672

[13] Qi, S. (2018) Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection. *International Conference on Learning Representations*.

[14] Audibert, J., Michiardi, P., Guyard, F., *et al.* (2020) Usad: Unsupervised Anomaly Detection on Multivariate Time Series. *Proceedings of the* 26*th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Virtual Event, CA, 6-10 July 2020, 3395-3404. https://doi.org/10.1145/3394486.3403392

[15] Siffer, A., Fouque, P.A., Termier, A., *et al*. (2017) Anomaly Detection in Streams with Extreme Value Theory. *Proceedings of the* 23*rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax, 13-17 August 2017, 1067-1075. https://doi.org/10.1145/3097983.3098144