计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(5):308−315 [doi: 10.15888/j.cnki.csa.009084]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于图自编码器的无监督多变量时间序列异常检测①

严盛辉[1,2], 陈志德[1,2]

[1](福建师范大学 计算机与网络空间安全学院, 福州 350007)
[2](福建省网络安全与密码技术重点实验室 (福建师范大学), 福州 350007)
通信作者: 陈志德, E-mail: zhidechen@fjnu.edu.cn

摘　要: 针对多变量时间序列复杂的时间相关性和高维度使得异常检测性能较差的问题, 以对抗训练框架为基础提出基于图自编码的无监督多变量时间序列异常检测模型. 首先, 将特征转换为嵌入向量来表示; 其次, 将划分好的时间序列结合嵌入向量转换为图结构数据; 然后, 用两个图自编码器模拟对抗训练重构数据样本; 最后, 根据测试数据在模型训练下的重构误差进行异常判定. 将提出的方法与 5 种基线异常检测方法进行比较. 实验结果表明, 提出的模型在测试数据集获得了最高的 *F*1 分数, 总体性能分 *F*1 分数比最新的异常检测模型 USAD 提高了 28.4%. 可见提出的模型有效提高异常检测性能.

关键词: 异常检测; 多变量时间序列; 对抗训练; 图自编码器; 重构

引用格式: 严盛辉,陈志德.基于图自编码器的无监督多变量时间序列异常检测.计算机系统应用,2023,32(5):308–315. http://www.c-s-a.org.cn/1003-3254/9084.html

GAE-based Unsupervised Anomaly Detection of Multivariable Time Series

YAN Sheng-Hui[1,2], CHEN Zhi-De[1,2]

[1](College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350007, China)
[2](Fujian Provincial Key Laboratory of Network Security and Cryptology (Fujian Normal University), Fuzhou 350007, China)

Abstract: The complex time correlation and high dimension of multivariable time series lead to poor anomaly detection performance. In view of this, an unsupervised anomaly detection model of multivariable time series based on graph autoencoders (GAEs) is proposed with the adversarial training framework as the basis. First, features are converted into embedded vectors. Secondly, the divided time series and embedded vectors are converted into graph-structured data. Then, two GAEs are used to simulate the adversarial training and reconstruct data samples. Finally, the anomaly is determined according to the reconstruction error of the test data under the model training. The proposed method is compared with five baseline anomaly detection methods. The experimental results show that the proposed model achieves the highest *F*1 score on the test data set, and the overall performance *F*1 score is 28.4% higher than that of the latest anomaly detection model, namely, USAD. Therefore, it can be seen that the proposed model can effectively improve the performance of anomaly detection.

Key words: anomaly detection; multivariable time series; adversarial training; graph autoencoder (GAE); reconstruction

1　引言

随着自动驾驶汽车、智能建筑、水处理和分配厂等信息物理系统中互连设备和传感器的快速增长, 物联网的出现进一步推动网络物理系统应用于各种任务,

① 基金项目: 国家自然科学基金 (62277010, 61841701); 福建省自然科学基金 (2021J011013)
收稿时间: 2022-11-01; 修改时间: 2022-11-29; 采用时间: 2022-12-11; csa 在线出版时间: 2023-03-03
CNKI 网络首发时间: 2023-03-07

越来越需要监控这些设备免受攻击, 对于电网、通信网络等关键基础设施尤为重要. 但随着传感器数据复杂性和维度的增加, 人们无法手动监视这些数据, 需要自动异常检测方法以快速检测高维数据中的异常.

由于传感器收集的数据中缺乏异常的标签, 同时异常往往是不可预测和多样化的, 通常把异常检测视为无监督学习问题. 近年来, 研究人员已经开发了许多经典的无监督方法, 包括基于统计过程控制的方法[1]、基于线性模型的方法[2] 和基于支持向量机的方法[3]. 如今的信息物理系统产生的多元时间序列是高度复杂的, 而且内在相关性是非线性, 然而这些方法只是简单的对传感器之间的关系进行建模, 仅能捕获线性关系.

因此, 研究人员开始利用基于深度学习的技术对高维数据进行异常检测, 从而开发出更智能和效益高的方法来识别异常[4]. 比如基于自编码器的异常检测方法[1] 和基于长短期记忆网络的异常检测方法[5]. 但是大多数的方法没有学习传感器彼此的关系, 在对有大量潜在相互关系的高维传感器数据进行建模时有困难. 近年来, 研究人员基于生成对抗网络的方法[6] 提出新的异常检测对抗训练框架, 但是生成对抗网络存在模式崩溃和不收敛的问题, 使得模型训练困难[7]. 图能很好地展现关系复杂的高维数据, Defferrard 等人提出图神经网络[8] 成功对图结构数据进行建模. 图神经网络以图结构数据作为输入, 若应用于多元时间序列的异常检测, 需要将时间序列中的复杂关系转换为图并与模型一起学习.

考虑到上述问题, 本文提出一个基于图自编码器的时间序列异常检测方法 GAE-AD, 该方法先学习传感器之间的关系图, 然后使用两个图自编码器进行 GAN 式对抗训练, 最后通过计算重构误差值进行异常检测. 本文主要贡献如下: (1) 本文提出了基于图自编码器的对抗训练框架的异常检测方法, 可以学习传感器之间依赖关系的图结构数据并结合自编码器和对抗训练的优势. (2) 对两个数据集进行了实验, 结果表明本方案与基线方法相比, 提升了对异常的检测准确率, 在 $F1$ 值和召回率两个异常检测指标上优于基线方法.

本文第 2 节介绍了相关工作. 第 3 节介绍本文提出的基于图自编码器的时间序列异常检测方法. 第 4 节在测试数据集上评估本文的方法. 第 5 节总结本文的工作.

2 相关工作

在许多领域中都会应用时间序列, 时间序列中的异常包含重要信息且会对整体产生重大影响. 异常检测方法主要有基于相似度、基于预测和基于重构等方法.

基于相似度的方法: 基于相似度的方法将远离大部分正常点的对象定义为异常值. 经典的方法包括基于聚 类的方法、基于距离的方法、基于密度的方法.

基于聚类的方法: 群集中样本点与其他样本点的距离、隶属性用来评估样本点的离群值, 比如 CBLOF[9], 将数据聚类以区分大小簇, 然后计算样本点与大簇的聚类中心点的距离, 作为异常得分.

基于距离的方法: 用群集中样本点到其他变量的距离作为异常值, 具有大的异常值的点被定义为异常. 比如 K-nearest neighbor 是一种常用的有监督学习算法[10], 计算测试样本与其最近 K 个样本的距离作为异常得分, 得分大于阈值的样本分类为异常点.

基于密度的方法: 用测试样本点在指定的区域内其他样本点的数量来定义密度, 在低密度区域的样本点分类为异常点. 比如 LOF[11], 将测试样本点周围样本点密度的平均值与其位置的密度之比定义为异常得分, 异常得分越大于 1, 表明测试样本点位置的密度相对于周围样本点越小, 它是异常点的可能性越大.

这些方法都是基于相似度的概念, 它们之间差异在于如何定义测试样本点与其他样本点的相似程度. 虽然基于相似度的方法简单, 易于推广. 但是在时间序列中, 不同的时刻具有前后相关性, 基于相似度的方法无法捕获这些相关性.

基于预测的方法: 该方法的思路是通过判断测试样本点的预测输入与其真实值之间的差异是否超过预设的阈值, 若超过则将该样本点识别为异常. 比如时序预测模型 ARIMA 方法[12], 该方法需要将时间序列转换为平稳、突变较少的序列, 使用前一时间段内的样本点来预测下一时刻的样本输入值. ARIMA 一共有 7 个参数, 选择合适的参数比较困难, 同时需要进行大量的参数估计, 导致 ARIMA 训练开销大. 基于迭代决策树 GBDT[13] 的异常检测, 通过梯度提升来对决策树迭代, 寻找最优的模型参数和学习率, 最终形成一个强学习器来预测输入值. 该方法缺点在于提取有效特征需要人工经验, 同时检测结果不稳定. 因为要使用前一个时间段来预测下一时刻的输入值, 所以基于预测的异常检测方法有滞后. 同时基于预测的方法容易受到异常点和周期性因素的影响, 适合平稳的时间序列.

基于重构的方法: 该方法通过构建一个编码-解码

网络执行重构任务, 该网络能正常重构正常样本, 而对异常样本的重构效果较差. 将测试样本的原始值和重构结果之间的差值定义为重构误差. 如果测试样本的重构误差值越大, 则该样本是异常的可能性越高. 基于自编码器的方法是最朴素的重构思想, 比如 LSTM-AE[14], 自编码器由编码器和解码器两部分组成, 编码器提取输入矢量的隐藏表示, 解码器根据隐藏表示通过与编码器相同的变换进行重构, 使重构结果与原始输入误差最小. 但是基于自编码器的神经网络结构的泛化性较强, 使得训练模型也能很好重构异常样本, 致使检测的误报率上升. 基于生成对抗网络的方法, 比如 MAD-GAN[15], 用两个长短期循环神经网络来替换生成器和判别器, 生成器的输入是测试样本的时间序列, 根据其生成假的时间序列并传递给判别器, 由判别器来区分生成序列和实际的训练序列. 基于 GAN 框架的时间序列异常检测模型需要的超参数较多, 同时异常也是在不断变化的, 需要一定的人工经验来纠正.

通过图的方法可以用边来表示特征之间和不同时刻的依赖关系, 根据构造的边来进行建模. 将传感器的时间序列转换为图结构进行异常检测, 可以更好地学习前后时刻和关联传感器之间的依赖关系. 本文的基于图自编码器的时间序列异常检测方法吸收了图表示依赖关系的特点, 同时利用自编码器对测试样本进行重构来检测异常.

图自编码器: 图自编码器 (GAE) 是将自编码器的编码器和解码器用两个图神经网络构造, 是一个可以半监督和无监督学习图结构数据的神经网络. Kipf 等人提出一种变分图自编码器[16], 该方法使用图卷积网络来构造编码器和解码器, 以图的邻接矩阵和节点信息作为输入, 学习图结构数据的隐藏向量表示. 变分图自编码器训练中使用图卷积网络, 训练开销会比较大. 图自编码器的优点是继承了无监督和半监督的特征, 编码和解码过程是对包含了深层次的拓扑结构信息和节点信息的图结构数据进行的, 有利于对图结构数据的信息传递、特征提取和表示.

在这项工作中, 将时间序列数据之间的关系转化为图结构, 引入两个图自编码器构成对抗训练结构进行重构, 并且用重构误差作为异常得分, 提高了异常检测的性能.

3 基于 GAE 的异常检测

将多变量数据序列的训练数据表示为 $D_{\text{train}}=\{D_{\text{train},1}, \cdots, D_{\text{train},n}\}$, 由于采用无监督的方法来学习异常检测模型, 所以训练数据由正常数据组成, 没有发生异常; 测试数据表示为 $D_{\text{test}}=\{D_{\text{test},1}, \cdots, D_{\text{test},m}\}$. 在时刻 t, $D_{\text{train},t}\in R^N$ 是一个 N 维向量, 当 N=1 时, 是单变量时间序列.

异常检测是一个二分类问题, 所以训练模型对于测试数据 D_{test} 的输出是二进制标签 $y(i)\in\{0, 1\}$, 若 $y(i)$=1, 表示时刻 i 出现了异常.

3.1 构建关系图

收集数据的传感器之间有着完全不同的特征, 但它们也可能以某种方式相互关联. 例如, 测量水位和水速的传感器, 二者之间的特征相差甚远, 但是水速的变化, 也会影响着水位的数据. 因此, 用以表示每个传感器的方式要能多维度捕捉传感器不同行为背后的联系.

因此, 为每个传感器引入一个能表示其特征的嵌入向量:

$$g_i\in\mathbb{R}^d,\ i\in\{1,2,\cdots,N\} \tag{1}$$

其中, g_i 表示第 i 个传感器在连续的 d 个时刻内数值所组成的向量.

嵌入向量 g_i 之间变化的相似性代表着传感器之间行为的相似性, 嵌入向量相似的传感器之间大概率有着高度相互关联. 这些嵌入向量将用于后续学习图的结构. 传感器关系如图 1 所示.

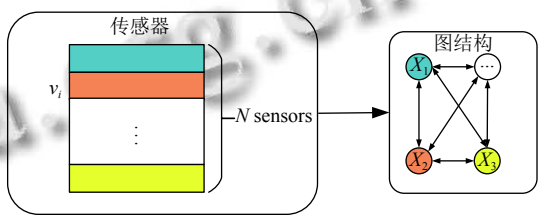


图 1 传感器关系图

3.2 学习图的结构

通过图结构来学习传感器之间的关联. 因为无向图是对称的, 无法表示传感器之间的不对称的依赖关系和因果关系. 所以本文将采用有向图连接特征, 展示不同传感器之间的依赖, 图的节点表示传感器, 节点之间的边代表它们的依赖关系. 使用邻接矩阵 A 来表示上述的有向图, 其中 A_{ij} 表示从节点 i 到节点 j 存在有向边.

根据图结构里存在的先验信息可以为每个传感器 i 创建一组关系 B_i, 即传感器 i 可能需要依赖的传感器:

$$B_i\subseteq\{1,2,\cdots,N\}-\{i\} \tag{2}$$

若没有先验信息, 传感器 i 的依赖关系包含除了自身外的所以传感器.

为了量化 B_i 中的依赖关系, 计算传感器 i 的嵌入向量与传感器 $j \in B_i$ 的嵌入向量之间的相似性:

$$e_{ji} = \frac{g_i^{\mathrm{T}} g_j}{\|g_i\| \cdot \|g_j\|}, j \in B_i \tag{3}$$

本文先计算 e_{ji}, 即传感器 i 的嵌入向量与传感器 j 之间的归一化点积.

综上所述, 给出构建图数据的算法 (算法 1). 步骤 1: 输入测试数据集所用的特征 F 和特征数量 n 构造候选关系, 第 i 个特征 F_i 的候选关系是除自身之外的 n–1 个特征的集合, 即 $\{F_1, \cdots, F_{i-1}, F_{i+1}, \cdots, F_n\}$. 步骤 2: 如果特征 F_i 是数据集的特征之一, 构造 F_i 与其候选关系的边. 步骤 3: 根据步骤 2 所构造边的特征数据计算图数据, 并对其归一化.

算法 1. 图数据构建

Input: 输入样本的特征和特征数 n
Output: 训练用的图数据

1. 根据特征和特征数 n 建立候选关系矩阵
2. 利用候选关系构造边
3. 根据步骤 2 中的构造好的边计算图数据并进行归一化

然后选择 k 个这样的归一化点积, k 是时间窗口大小, 用来模拟当前时间点和之前时间点, 以此将时间序列划分成子序列. 给定长度为 k 的时间窗口, 组成邻接矩阵:

$$A_{ji} = \{j \in \{e_{ki} : k \in B_i\}\} \tag{4}$$

后续模型的训练将用到该邻接矩阵.

3.3 训练模型

图自编码器 (GAE)[17] 是用于无监督学习、高效编码的神经网络, 结合了编码器和解码器, 其目的在于重构编码器的输入, 利用反向传播, 使得解码器的输出目标和输入相同.

编码器获取输入 X 并将其映射到一组潜在变量 Z, 解码器将潜在变量 Z 映射会输入空间作为重构结果 R. 原始输入 X 与重构结果 R 之间的误差称为重构误差. 本文的目的是最小化该重构误差. 重构误差定义如下:

$$\begin{cases} f: X \to Z \\ g: Z \to R \\ f, g = \arg\min \|X - g[f(X)]\|^2 \end{cases} \tag{5}$$

基于图自编码器的异常检测 GAE-AD 使用重构误差作为异常分数, 分数超过阈值的点被分类为异常. 如果异常非常小, 与正常数据比较接近, 则重构误差会很小, 导致该异常无法检测. 这是因为自编码器旨在重构输入数据以接近正态性. 所以, 图自编码器在训练之前应当识别输入的数据是否是正常数据. 而由于在训练时, 使用的数据不包含异常, 所以, 在测试时, 图自编码器能很好地重构正常数据, 而由于训练过程中不包含异常数据, 对异常数据的重构效果不好, 所得到的重构误差较大.

生成对抗网络 (GAN) 是判断输入是否为异常的一种方法, GAN 是基于两个网络之间极大极小博弈. 一个网络是生成网络, 旨在生成样本使判别网络无法区分其真实性; 另一个是判别网络, 目标是尽量区分数据源自于真实数据还是生成网络生成的. 基于生成对抗网络的异常检测使用正常数据训练, 用判别网络作异常检测器, 但是生成对抗网络会出现模式崩溃和不收敛的问题, 这是由于生成网络和判别网络之间的不平衡[8].

本文方法 GAE-AD 分为两阶段对抗训练的图自编码器结构. 这是因为自编码器在对抗训练中能获得稳定性, 解决了生成对抗网络的模式崩溃和不收敛的问题, 同时也能对输入数据进行良好的重构.

本文框架 GAE-AD 由两个编码器网络 encoder1 (E_1)、encoder2 (E_2) 和两个解码器网络 decoder1 (D_1)、decoder2 (D_2) 组成, 构成了两个图自编码器 GAE_1 和 GAE_2, 如图 2.

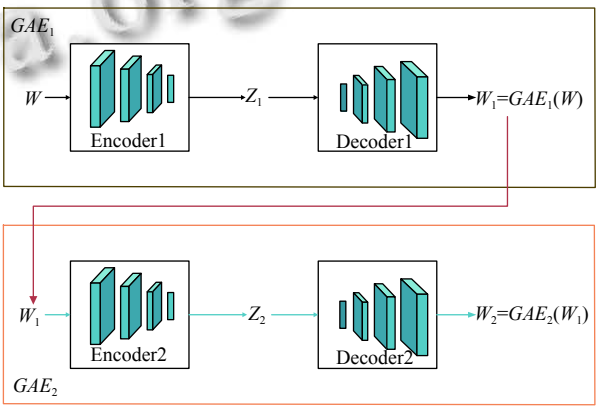


图 2 GAE-AD 模型

首先对两个图自编码器进行训练 (算法 2), 学习重构正常的输入 W; 然后, 两个图自编码器以对抗的方式进行训练, 如同生成对抗网络极大极小博弈, GAE_1 试图欺骗 GAE_2, 而 GAE_2 要尽量区分数据是输入的 W 还

是来自于 GAE_1 重构的.

算法 2. 训练算法

Input: 输入样本 $W=W_1, \cdots, W_n$, 训练次数 N
Output: 训练好的 GAE_1 和 GAE_2

1. 初始化 E_1、E_2、D_1、D_2, $n\leftarrow 1$
2. **while** $n!=N$ **do**
3. for $i=1$ to T do
4. $Z_i\leftarrow E_1(W_i)$
5. $W_i^1\leftarrow D_1(Z_1)$
6. $Z_2\leftarrow E_2(W_i^1)$
7. $W_i^2\leftarrow D_2(Z_2)$
8. $loss_{GAE_1}\leftarrow \frac{1}{n}\|W_i-W_i^1\|_2+(1-\frac{1}{n})\|W_i-W_i^2\|_2$
9. 使用 $loss_{GAE_1}$、$loss_{GAE_2}$ 更新 E_1、E_2、D_1、D_2 权重
10. end for
11. $n\leftarrow n+1$
12. end while

第 1 个阶段: 图自编码器训练. 在这个阶段, 输入数据 W 有编码器 E_1 压缩到潜在空间 Z, 然后由解码器 D_1 进行重构. 第 1 阶段的图自编码器训练如下:

$$GAE_1(W)=D_1(E_1(W)) \tag{6}$$

第 2 个阶段: 对抗训练. 这个阶段的目标是训练 GAE_2 区分数据是真实数据还是来源于 GAE_1, 同时 GAE_1 继续学习以欺骗 GAE_2. GAE_1 的输出通过编码器 E_2 压缩到潜在空间 Z, 再由 GAE_2 的解码器 D_2 重构. 使用生成对抗网络的训练方法, 生成器的目标是最小化输入 W 和 GAE_2 重构结果的不同, 鉴别器则旨在最大化二者之间的差异以实现异常检测. 第 2 阶段训练的如下:

$$GAE_2(W)=D_2(E_2(W)) \tag{7}$$

综上, 两个阶段类似生成对抗网络的极大极小博弈, 公式描述如下:

$$\min_{GAE_1}\max_{GAE_2}V(GAE_1,GAE_2)=\min_{GAE_1}\max_{GAE_2}\|W-GAE_2(GAE_1(W))\|_2 \tag{8}$$

第 1 个阶段的训练目标是最小化重构误差 $\|W_i-W_i^1\|_2$, 来学习数据潜在特征, 第 2 个阶段的训练目标是将重构误差 $\|W_i-W_i^2\|_2$ 降至最低. 对两个阶段设置了重构误差的权重比例, 随着训练迭代而变化. 最终将两个阶段的训练损失结合起来所得到的损失函数式 (9) 所示:

$$loss_{GAE_1}\leftarrow\frac{1}{n}\|W_i-W_i^1\|_2+\left(1-\frac{1}{n}\right)\|W_i-W_i^2\|_2 \tag{9}$$

在检测阶段 (算法 3), 异常得分定义为:

$$score(W')=\frac{1}{2}\|W'-GAE_1(W')\|_2+\frac{1}{2}\|W'-GAE_2(GAE_1(W'))\|_2 \tag{10}$$

如果时刻 i 的异常分数高于阈值 threshold, 则该时刻被分类为异常.

算法 3. 异常检测算法

Input: 输入测试数据 $W'=W_1',\cdots,W_m'$, 阈值 λ
Output: 分类结果: $y_1, \cdots, y_m$

1. 初始化异常分数 $score\leftarrow[]$
2. for $i=1$ to m do
3. $W_i'^1\leftarrow D_1(E_1(W_i'))$
4. $W_i'^2\leftarrow D_2(E_2(W_i'^1))$
5. $score$.append$(0.5\|W_i'-W_i'^1\|_2+0.5\|W_i'-W_i'^2\|_2)$
6. if $score\geqslant\lambda$ then
7. $y_i\leftarrow 1$
8. else
9. $y_i\leftarrow 0$
10. end if
11. end for

4 实验

本文使用的数据集是基于水处理物理试验台系统的两个传感器数据集: SWaT 和 WADI, 均是由新加坡科技设计大学的 iTrust 机构采集并开源. 安全水处理数据集 (SWaT) 来自新加坡公共事业委员会[16]. 它代表了小规模的信息物理系统, 集成了数据和物理实体来控制和监控系统行为. 配水数据集 (WADI) 是 SWat 的延伸, 由一个大量配水管道组成的配水系统[18], 是个更完整、更现实的水处理、储存和分配网络.

如表 1 所示, SWaT 数据集有 51 个特征, 总的样本数目为 946 719, 其中正常工作下的训练样本有 568 031 个, 带有异常的测试样本数为 378 688, 有 11.97% 的异常. WADI 有 127 个特征, 总共有 577 658 个样本, 其中正常工作下的训练样本有 346 594 个, 带有异常的测试样本数为 231 064, 异常占比 5.99%.

表 1 数据集统计信息

数据集	样本数目	训练	测试	异常占比 (%)	特征
SWaT	946 719	568 031	378 688	11.97	51
WADI	577 658	346 594	231 064	5.99	127

这两个数据集包含两周的正常运行数据, 这些数据作为训练. 在接下来的几天里, 在不同的时间间隔进

行了受控的物理攻击以模拟异常情况, 这段时间内的数据用作测试数据.

本文使用 Python 3.8 作为编程基础, PyTorch 1.10.0、cuda 11.1 实现和测试. 使用 Windows 10、Intel(R) Core(TM) i7-10750H CPU、RTX3060、16 GB 内存的计算机运行程序. 实验选取的滑动窗口 k=12, 模型的训练迭代次数 N=50, 批大小 batch=8, 模块的编码网络和解码网络均是图神经网络.

4.1 基准模型

本文选择了 5 个异常检测模型作为基准模型, 与本文提出的基于图自编码器的异常检测模型 GAE-AD 进行比较.

LSTM-VAE[19]: 用长短期记忆网络替换变分自编码器的前馈神经网络以结合二者, 对多维信号的底层分布进行建模, 并使用预期的分布信息重建原始的时间序列, 通过重构误差检测异常.

DAGMM[20]: 深度自编码高斯模型将自编码器和高斯混合模型联合起来的无监督异常检测模型, 网络结构分为压缩网络和估计网络, 压缩网络将样本映射到异常检测信息的低维空间, 估计网络评估低维空间中样本的高斯混合建模, 最后使用重构误差判定异常.

Isolation Forest[21]: IF 是基于决策树的算法, 在给定的数据集中构建 iTrees 的集合, 经过重复划分递归后, 在 iTrees 上具有较短平均路径长度的样本被归为异常, 较长平均路径长度则表示为正常.

AE[22]: 使用自编码器的编码器和解码器来重构数据样本, 将重构误差作为异常分数, 如果测试样本的异常分数高于预设的阈值, 则判断为异常.

USAD[23]: 基于重构的无监督方法, 由 1 个编码器和 2 个解码器构成, 在训练过程中学习如何放大异常样本输入的重构误差, 利用重构误差识别异常.

4.2 评估指标

准确率 (P)、召回率 (R) 和 $F1$ 值用于评估异常检测性能. TP 表示异常样本经过模型分类为异常的个数. FN 表示异常样本经过模型分类为正常的个数. FP 表示正常样本经过模型分类为异常的个数.

$$\begin{cases} P = \dfrac{TP}{TP+FP} \\ R = \dfrac{TP}{TP+FN} \\ F1 = \dfrac{2P\times R}{P+R} \end{cases} \tag{11}$$

其中, P 为检测的准确率, 表示检测到的异常样本的百分比; R 为召回率, 表示正确识别出来的异常样本的百分比, R 越高表示漏报的异常越少; $F1$ 值兼顾准确率和召回率, 是评估模型异常检测性能的主要指标.

4.3 结果分析

为了展示本文模型 GAE-AD 的整体性能, 对比了 5 个基准模型在多变量时间序列 SWaT 和 WADI 数据集上的检测性能, 比较结果如表 2、图 3、图 4 所示. 由表 2 可知, 本文方案在两个测试数据集上的召回率和 $F1$ 值均为最高, 异常检测整体性能最好.

表 2 模型性能 (%)

方法	SWaT			WADI			平均$F1$
	P	R	$F1$	P	R	$F1$	
DAGMM	27.46	69.52	39.37	54.44	26.99	35.79	37.63
AE	72.63	52.63	61.03	39.70	32.20	35.56	48.29
IF	95.12	58.84	72.71	29.92	15.83	20.71	46.71
LSTM-VAE	96.24	59.91	73.85	**87.79**	14.45	24.82	49.33
USAD	**98.70**	74.02	84.60	64.51	32.20	42.96	63.78
GAE-AD	86.45	**87.36**	**86.90**	68.96	**86.95**	**76.91**	**81.90**

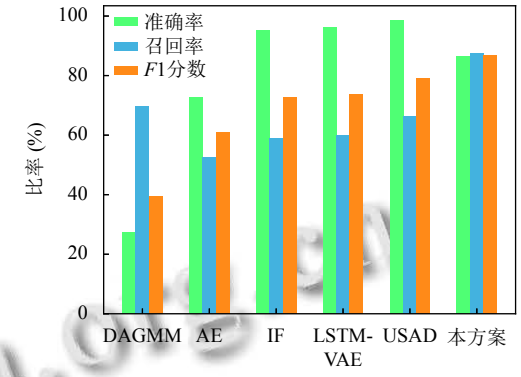


图 3 SWaT 上模型性能

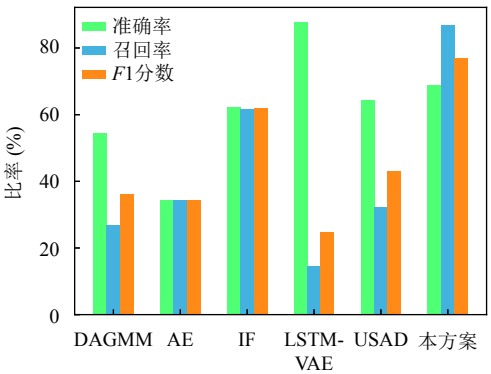


图 4 WADI 上模型性能

DAGMM 性能最差的原因是它没有利用时间序列中的时间信息, 对于时间序列, 这些信息十分重要, 不

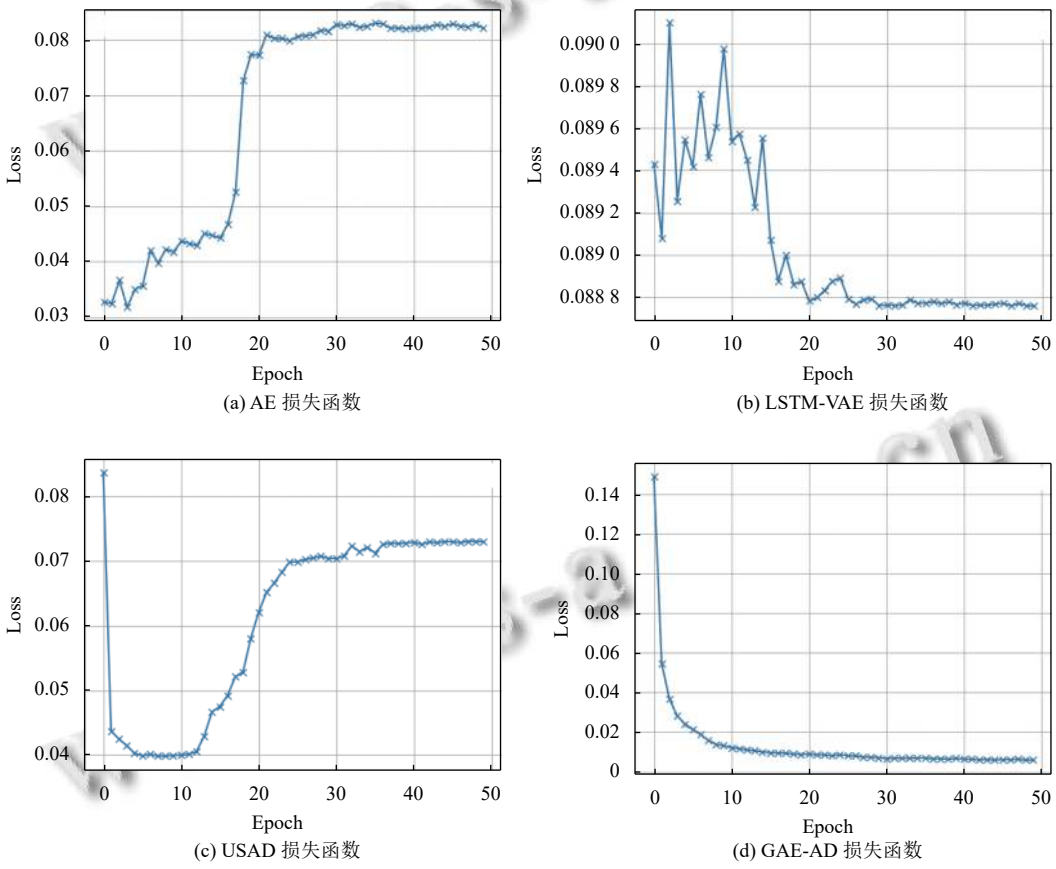
同时刻的观测数据是相关的, 历史数据有助于重构当前时刻的样本. 在本文模型 GAE-AD 中, 输入是一系列观测数据, 包含了时间序列的时间信息和关系, 使用图结构数据可以学习这些信息和关系.

IF 方法适用于低维的数据. 在每次切分数据空间时, IF 是随机选取一个维度, 如果数据维度高, 在构建好树后, 会有大量的维度信息没有使用. 本方案基于图自编码器这一深度学习方法, 可以提取高维时间序列数据的特征.

AE、USAD 和 LSTM-VAE 都能保留时间序列的时间信息, 能很好地重构输入样本, 使得这 3 种方法无法检测接近正常样本的异常. GAE-AD 通过基于图自编码器的对抗训练弥补了这一缺点, 能更好地识别出异常样本, 使召回率相对于 5 个基准模型有了大幅提升.

从表 1 可以看出, WADI 的特征维度远高于 SWaT, 因此 WADI 也比 SWaT 也更加不平衡, 所以在反映异常检测整体性能的 $F1$ 值上, 6 个模型在 WADI 数据集上的性能相对于 SWaT 数据集出现了下降.

图 5 展示了 AE、LSTM-VAE、USAD、GAE-AD 这 4 个模型的损失函数曲线, 横坐标为迭代次数, 纵坐标为损失函数值. 从图 5 可以看出, 训练初期梯度更新信息变化较大, 但是随着迭代次数增加, 损失值趋于稳定并且保持在一定范围内振荡. 相比其他 3 个模型, 本文提出的 GAE-AD 收敛速度最快, 最先达到稳定状态, 并且它的损失值是最小的. 说明 GAE-AD 与其他 3 个模型相比, 具有损失小、收敛快的优势.



(a) AE 损失函数 (b) LSTM-VAE 损失函数

(c) USAD 损失函数 (d) GAE-AD 损失函数

图 5 不同模型损失函数曲线

5 结语

本文对多变量时间序列进行异常检测, 提出基于图自编码的异常检测模型 GAE-AD. 通过实验表明其提高了异常检测的性能. 本文考虑到时间序列相关性引入了图自编码器; 同时针对生成对抗网络的模式崩溃和不收敛的问题, 引入了两个自编码器组成对抗训练架构; 将传感器收集的时间序列数据转换为图结构数据并且划分成多个子序列, 能更好地学习前后时刻之间的关系, 有助于发现时间序列中的上下文异常和提升异常检测的召回率, 通过实验表明其提升了异常检测的性能. 在多变量时间序列 SWaT 和 WADI 数据集上评估本文模型, 并且与其他基准模型对比, 发现

GAE-AD 模型获得了最好的总体性能 (最高 *F*1 分数).

通过对比 DAGMM、AE、IF、LSTM-VAE、USAD 这 5 种异常检测模型在 SWaT 和 WADI 两个多变量时间序列数据集上异常检测性能可知, 基于图自编码器的多维时间序列异常检测方法 GAE-AD 能更好学习时间序列的前后时刻关系, 放大学习异常样本的重构误差, 更好地区分正常和异常样本以提升召回率. 但是本方案的准确率较低, 如何提升本文方法异常检测准确率是后续研究方向.

参考文献

1 Aggarwal CC. Outlier Analysis. 2nd ed., New York: Springer, 2017.

2 Shyu ML, Chen SC, Sarinnapakorn K, *et al*. A novel anomaly detection scheme based on principal component classifier. Proceedings of the Foundations and New Directions of Data Mining Workshop. IEEE, 2003. 172–179.

3 Schölkopf B, Platt JC, Shawe-Taylor J, *et al*. Estimating the support of a high-dimensional distribution. Neural Computation, 2001, 13(7): 1443–1471. [doi: 10.1162/089976601750264965]

4 Kwon D, Kim H, Kim J, *et al*. A survey of deep learning-based network anomaly detection. Cluster Computing, 2019, 22(1): 949–961.

5 Qin Y, Song DJ, Cheng HF, *et al*. A dual-stage attention-based recurrent neural network for time series prediction. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: AAAI Press, 2017. 2627–2633.

6 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2672–2680.

7 Arjovsky M, Bottou L. Towards principled methods for training generative adversarial networks. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017. 1050.

8 Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 3844–3852.

9 He ZY, Xu XF, Deng SC. Discovering cluster-based local outliers. Pattern Recognition Letters, 2003, 24(9–10): 1641–1650.

10 冯贵兰, 周文刚. 基于 Spark 平台的并行 KNN 异常检测算法. 计算机科学, 2018, 45(S2): 349–352, 366. [doi: 10.11896/j.issn.1002-137X.2018.11A.071]

11 杭菲璐, 郭威, 陈何雄, 等. 基于 iForest 和 LOF 的流量异常检测. 计算机应用研究, 2022, 39(10): 3119–3123. [doi: 10.19734/j.issn.1001-3695.2022.03.0121]

12 Moayedi HZ, Masnadi-Shirazi MA. Arima model for network traffic prediction and anomaly detection. Proceedings of 2008 International Symposium on Information Technology. Kuala Lumpur: IEEE, 2008. 1–6.

13 Wen ZY, He BS, Kotagiri R, *et al*. Efficient gradient boosted decision tree training on GPUs. Proceedings of 2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Vancouver: IEEE, 2018. 234–243.

14 陈磊, 秦凯, 郝矿荣. 基于集成 LSTM-AE 的时间序列异常检测方法. 华中科技大学学报 (自然科学版), 2021, 49(11): 35–40. [doi: 10.13245/j.hust.211107]

15 Li D, Chen DC, Jin BH, *et al*. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. Proceedings of the 28th International Conference on Artificial Neural Networks. Munich: Springer, 2019. 703–716.

16 Kipf TN, Welling M. Variational graph auto-encoders. arXiv:1611.07308, 2016.

17 吴博, 梁循, 张树森, 等. 图神经网络前沿进展与应用. 计算机学报, 2022, 45(1): 35–68. [doi: 10.11897/SP.J.1016.2022.00035]

18 Ahmed CM, Palleti VR, Mathur AP. WADI: A water distribution testbed for research in the design of secure cyber physical systems. Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks. Pittsburgh: ACM, 2017. 25–28.

19 Park D, Hoshi Y, Kemp CC. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. IEEE Robotics and Automation Letters, 2018, 3(3): 1544–1551. [doi: 10.1109/LRA.2018.2801475]

20 Zong B, Song Q, Min MR, *et al*. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018.

21 Liu FT, Ting KM, Zhou ZH. Isolation forest. Proceedings of the 8th IEEE International Conference on Data Mining. Pisa: IEEE, 2008. 413–422.

22 Chow JK, Su Z, Wu J, *et al*. Anomaly detection of defects on concrete structures with the convolutional autoencoder. Advanced Engineering Informatics, 2020, 45: 101105. [doi: 10.1016/j.aei.2020.101105]

23 Audibert J, Michiardi P, Guyard F, *et al*. USAD: Unsupervised anomaly detection on multivariate time series. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2020. 3395–3404.

(校对责编: 孙君艳)