

基于 GAT-AGRU 的多元时序数据异常检测^{*}杨晨龙¹ 孙 晔² 刘晓悦¹

(1. 华北理工大学电气工程学院 唐山 063210; 2. 中国信息通信研究院 北京 100191)

摘要: 在复杂工业系统因传感器数量急剧增加,产生了高维噪声和随机干扰,严重影响多元时间序列的数据连续性和控制精度。然而现有对多元时间序列存在随时间变化的时序不一致性、空间矢量的偏差性以及时空图模型的冗余度等问题。本文提出了一种新的多元时间序列异常检测方法 STGAD。首先,从高分辨率的粒度级别上,引入门控机制改进多尺度卷积网络,控制特征间的信息交互过程。然后,设计了两种图结构,剔除了冗余的时空依赖关系,利用 GAT 有效地学习时空相关性。此外,提出了一种基于注意力机制的 GRU 模块,来捕获变量在不同时间窗口上的重要性。最后,联合优化预测和重构的模块。在三个公开数据集上进行广泛的实验,结果表明所提出模型的平均 F1 分数高于 0.94,在高维数据集上明显优于其他基准模型。

关键词: 多元时间序列;异常检测;图注意力网络;门控机制;注意力机制

中图分类号: TN911.71 **文献标识码:** A **国家标准学科分类代码:** 510.4030

Anomaly detection of multivariate time-series based on GAT-AGRU

Yang Chenlong¹ Sun Ye² Liu Xiaoyue¹

(1. College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China;

2. China Academy of Information and Communications Technology, Beijing 100191, China)

Abstract: In complex industrial systems due to the dramatic increase in the number of sensors, high-dimensional noise and random disturbances are generated, which seriously affect the data continuity and control accuracy of multivariate time series. However, the existing pairs of multivariate time series have the problems of temporal inconsistency over time, deviation of space vectors, and redundancy of spatio-temporal graphical models. In this paper, a new multivariate time series anomaly detection method STGAD is proposed. First, a gating mechanism is introduced to improve the multiscale convolutional network from a high-resolution granularity level to control the process of information interaction between features. Then, two graph structures are designed to eliminate redundant spatio-temporal dependencies, enabling GAT to effectively learn spatio-temporal correlations. In addition, an attention mechanism-based GRU module is proposed to capture the importance of variables over different time windows. Finally, modules for joint optimization prediction and reconstruction. Extensive experiments on three publicly available datasets show that the average F1-score of the proposed model is higher than 0.94, which significantly outperforms other benchmark models on high-dimensional datasets.

Keywords: multivariate time-series; anomaly detection; graph attention network; gated mechanism; attention mechanism

0 引言

工业控制系统的传感器设备迅速增加,产生了大量的时间序列数据^[1]。这些数据表现出复杂的非线性拓扑关系和时间模式,并且容易受到相应的异常和网络攻击^[2]。时间序列异常检测有着广泛的应用,例如智能电网^[3]、水处理和配网^[4]、交通系统和自动驾驶^[5]。通过持续监测基本控

制或指标,可以及时警报潜在的故障,因此,一个高效准确的异常检测系统具有很大的研究价值。

由于异常的缺乏和多样性,这使得数据标记具有挑战性,大多数方法都遵循无监督学习。近年来,基于深度学习的技术在异常检测领域广泛应用,现有方法大致可以分为两类:基于预测的模型和基于重构的模型。例如,基于循环神经网络(recurrent neural network, RNN)^[6]的预测方法

能够捕获时间依赖关系,利用预测与真实值之间的偏差来检测异常。基于重构的自动编码器(autoencoder, AE)^[7]和生成对抗网络(generative adversarial network, GAN)^[8]等方法可以通过原始样本和重构样本之间的差值获得重构误差,有效地拟合数据进行异常检测。然而,这些方法并没有明确地学习传感器之间的拓扑结构,现实世界中,大多数数据都是从非欧氏空间中生成的。许多深度学习方法在处理这些数据方面表现不佳。

近年来,图神经网络(graph neural network, GNN)因其有效地建模图结构数据而愈受欢迎,在捕获空间相关性取得了很大进展^[9]。其中图注意力网络受到更多的关注,相比于图卷积网络(graph convolutional network, GCN)^[10],图注意力网络(graph attention network, GAT)^[11]通过注意力的引入能够更好地学习节点之间的关系,允许每个节点仅关注与其相关的邻居节点,从而提升图结构中的表达能力。一些方法也成功地应用在了多元时序数据异常检测方面。例如,MTAD-GAT^[12]从时间和空间两个角度建模图结构,利用并行的 GAT 全面地学习的时空相关性,同时从预测和重构的角度计算误差,取得了很好的效果。虽然这种结构可以对传感器和时间的相关性进行显示建模,但使用了两个完全图,存在信息冗余。这些方法都没有充分考虑时空特性,对图结构的设计仍有改进空间。而且只关注了变量间的相关性,却忽略了变量在不同时段的重要性。

基于上述问题和挑战,本文提出了一种基于 GAT-AGRU 的多元时序数据异常检测方法。首先,提出了基于门控机制改进的多尺度卷积网络,传递相关信息,提取细粒度层次的特征。然后,从时空角度设计了两种不同的图结构,以减少信息冗余,降低模型的计算复杂度。此外,使用注意力机制改进门控循环单元(gate recurrent unit, GRU),学习变量在不同时间窗口中的重要性。最后,结合

预测和重构模块进行异常评分。在 3 个公共数据集上进行了 F1 性能指标的评估,并进行了消融实验,证明了该方法的有效性。

1 STGAD 模型

1.1 问题定义

多元时间序列是与时间戳相关的观测值或数据点的集合,记为 $x_t = \{x_1, x_2, \dots, x_n\}$, $x_t \in R^{n \times m}$ 。其中 n 代表时间戳的最大值, m 代表特征数量。由于长时间序列包含大量数据点,所以采用固定的时间窗口作为输入,一个时间窗口记为 $w_t = \{x_{t-k+1}, \dots, x_t\}$, $w_t \in R^{k \times m}$ 。其中 k 表示时间窗口大小。检测任务是根据过去和当前的观测值预测下一时刻的值,输出向量是 $y_t = \{y_1, y_2, \dots, y_n\}$, 其中 $y_t \in \{0, 1\}$ 。当 $y_t = 1$ 时,表示观测值 x_t 异常,反之正常。

1.2 整体框架

STGAD 整体框架如图 1 所示,它主要包括以下 5 个部分:

1)多尺度门控卷积网络(GRes2Net)。首先使用多尺度卷积网络(Res2Net)^[13]提取多尺度特征图的局部特征,并在特征图的残差连接中加入门控机制,选择性地融合特征图,增强泛化能力。

2)图结构。空间上,利用 top-k 法构建了一个有向的稀疏空间图来建模传感器间的相关性,减少计算冗余。时间上,使用上三角矩阵构建了一个有向加权的时间导向图来建模时间相关性,精确地表达时间特点。

3)图注意力。使用两个并行的 GAT 层学习空间图和时间图,捕获时空相关性。

4)基于注意力机制的 GRU。学习变量在每个时间段基于时间依赖性的重要性。

5)异常得分。联合优化预测和重构模块,共同计算异常评分。

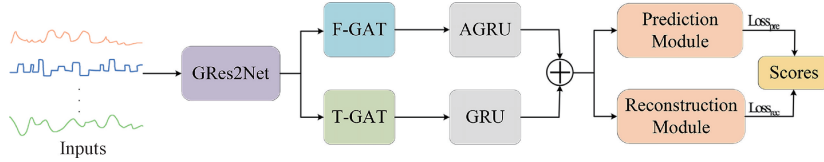


图 1 STGAD 整体架构

Fig. 1 The overall architecture of STGAD

1.3 多尺度门控卷积网络(GRes2Net)

卷积神经网络(convolutional neural network, CNN)擅长提取时间窗口内的局部特征。但是它提取的是固定大小上下文的特征。如图 2 所示,本文采用一种多尺度处理方法 GRes2Net,可以从细粒度级别提取局部特征。首先使用 1×1 卷积层扩展特征图通道,平均划分为 S 个特征图,每组特征图记为 x_i , 其中 $i \in \{1, 2, \dots, s\}$ 。然后使用 3×3 卷积层提取特征图的特征,将前一个输出的特征图和当前特征图融合发送到下一个卷积层提取特征。最

后将所有特征图连接起来,发送到 1×1 卷积层压缩操作,以融合所有特征图的信息。Res2Net 通过残差连接直接将前面的特征信息全部送给下一组,忽略了特征映射之间的相关性。通过在残差连接中加入门控机制来控制不同特征图之间的信息交互。输出 y_i 的表达式为:

$$y_i = \begin{cases} x_i, & i = 1 \\ f_i(x_i), & i = 2 \\ f_i(x_i + g_i \cdot y_{i-1}), & 2 < i \leq s \end{cases} \quad (1)$$

$$g_i = \tanh(f(\text{concat}(f(x_i), f(y_{i-1})))) \quad (2)$$

其中, g_i 表示门控选择操作, $f_i(\cdot)$ 表示卷积操作, 用前一个输出特征图和下一个特征图计算 g_i 的值, 用于控制残差连接的信息交互。

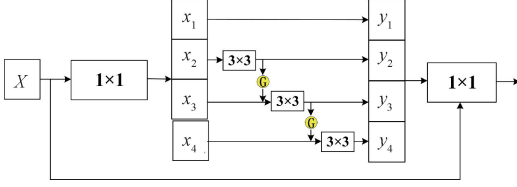


图 2 多尺度门控卷积

Fig. 2 Multi-scale gated convolution

1.4 图结构

1) 空间图

由于传感器之间的关系未必是相互的, 需要采用有向图来表示。同时冗余增加了计算复杂度和过拟合的风险, 因此通过 top-K 进行节点的稀疏化。如图 3 所示, 定义图中的每个节点代表一个传感器, 即 $v_i^M = \{v_1, v_2, \dots, v_m\}$, $v_i^M \in R^k$, 其中, M 表示空间层面, m 表示传感器的数量, k 是图节点的维度(即时间窗口大小)。节点之间相似度表示为:

$$e_{ij} = \text{sim}(v_i^M, v_j^M) = \frac{(v_i^M)^T v_j^M}{\|v_i^M\| \cdot \|v_j^M\|} \quad (3)$$

其中, $\text{sim}(\cdot)$ 表示余弦相似度计算, e_{ij} 表示节点 i 和节点 j 之间的相似度。然后选择前 K 个相似度值作为索引, 构建邻接矩阵为:

$$\mathbf{A}_{ij} = 1\{j \in \text{TopK}(e_{iK} \mid K \in C_i)\} \quad (4)$$

其中, $\mathbf{A}_{ij}$ 表示从节点 j 到节点 i 存在一条有向边, 选取与节点 i 相关的前 K 个节点, 令 $\mathbf{A}_{ij} = 1$, 否则 $\mathbf{A}_{ij} = 0$ 。

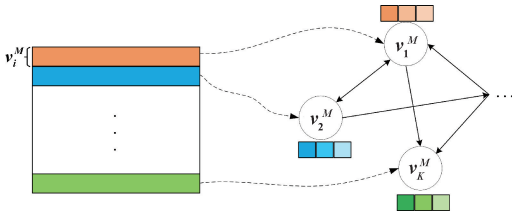


图 3 稀疏空间图

Fig. 3 Sparse spatial graph

2) 时间图

时间具有不可逆性, 为了防止后面的时间戳影响前面的时间戳, 按照时间顺序构建有向边。并且每个时间戳之间的关系不是相同的, 越近的相关性越大。随着时间流逝, 远处的时间影响逐渐变小, 因此赋予每个时间戳权重以表示它们之间的相关性。如图 4 所示, 定义每个节点代表一个时间戳, 即 $v_i^T = \{v_1, v_2, \dots, v_k\}$, $v_i^T \in R^m$, 其中 T 表示时间层面, k 表示时间窗口大小, m 是图节点的维度(即传感器的数量)。为了定义边的方向, 使用上邻接矩阵形式表示:

$$\mathbf{B}_{ij} = \begin{cases} 1, & i \leq j \\ 0, & i > j \end{cases} \quad (5)$$

其中, $\mathbf{B}_{ij} = 1$ 表示时间流方向上节点 i 到节点 j 有一条边。接下来将有向边赋予权重, 越近的时间戳影响力越大。

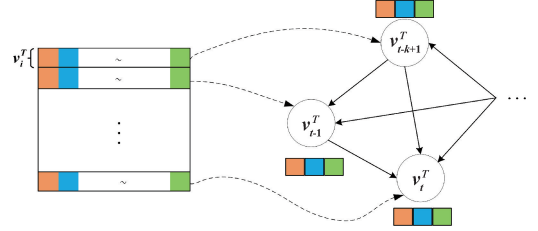


图 4 时间导向图

Fig. 4 Time-oriented graph

加权邻接矩阵定义为:

$$\mathbf{B}_{ij}^w = \begin{cases} \log_e(k - (j - i) + 1), & i < j \\ 1, & i = j \\ 0, & i > j \end{cases} \quad (6)$$

1.5 图注意力

GAT 能够根据图结构聚合相邻节点的信息, 并且自适应学习每个节点在不同邻居节点上的重要性。

1) 面向特征的 GAT 层

使用稀疏特征图 $\mathbf{A}_{ij}$ 作为输入来学习传感器之间的相关性, F-GAT 层每个节点的输出表示如下:

$$\tilde{h}_i^i = \sigma\left(\sum_{j \in C(i)} \alpha_{ij}^M v_j^M\right) \quad (7)$$

节点 i 与节点 j 之间的注意力得分 α_{ij}^M 计算公式为:

$$\alpha_{ij}^M = \text{softmax}(e_{ij}^M) = \frac{\exp(e_{ij}^M)}{\sum_{K \in C(i)} \exp(e_{iK}^M)} \quad (8)$$

$$e_{ij}^M = \mathbf{w}_M^T \cdot \text{LeakyReLU}(\mathbf{W}_M \cdot [v_i^M \parallel v_j^M]) \quad (9)$$

其中, σ 是一个非线性激活函数, $[\parallel]$ 是连接操作, $\mathbf{W}_M \in R^{2k \times 2k}$ 是一个可学习的权重参数矩阵, $\mathbf{w}_M \in R^{2k}$ 是注意力机制的可学习的参数向量, e_{ij}^M 是节点 i 与节点 j 的注意力系数, $C(i) = \{j \mid \mathbf{A}_{ij} = 1\}$ 是与节点 i 相似度排在前 K 的邻居节点集合。

2) 面向时间的 GAT 层

使用时间导向图 $\mathbf{B}_{ij}^w$ 作为输入来学习时间戳之间的依赖, T-GAT 层每个节点的输出表示如下:

$$h_i^T = \sigma\left(\sum_{j \in N(i)} \alpha_{ij}^T v_j^T\right) \quad (10)$$

节点 i 与节点 j 之间的注意力得分 α_{ij}^T 计算公式为:

$$\alpha_{ij}^T = \text{softmax}(e_{ij}^T) = \frac{\exp(e_{ij}^T)}{\sum_{k \in N(i)} \exp(e_{ik}^T)} \quad (11)$$

$$e_{ij}^T = \mathbf{w}_T^T \cdot \text{LeakyReLU}(\mathbf{W}_T \cdot [v_i^T \parallel v_j^T]) \quad (12)$$

其中, $\mathbf{W}_T \in R^{2m \times 2m}$ 是一个可学习的权重参数矩阵, $\mathbf{w}_T \in R^{2m}$ 是注意力机制的可学习的参数向量, e_{ij}^T 是节点 i 与节点 j 的注意力系数, $N(i) = \{j \mid i \geq j\}$ 是节点 i 的邻

居节点集合。

1.6 基于注意力机制的 GRU

虽然时空图已经建模变量间的相关性,但异常检测中每个变量重要性不是固定的。变量在不同时间窗口上的重要性可能不同,因此在时间维度上捕获变量的重要性是至关重要的。为此,本文提出一种基于注意力机制的 GRU 来学习每个变量在不同时间窗口中的重要性。之前的工作^[14]已经证明了 GRU 可以成功地捕获时间依赖性。基于注意力机制的 GRU 更简单高效,它可以大大减少训练时间,实现加权变量的功能。如图 5 所示,在每个 GRU 单元中,本文以 F-GAT 的输出 $\tilde{h}_i$ 作为输入,通过线性层和 softmax 层获得当前时间窗口内变量的权重。然后使用加权变量作为 GRU 单元的输入,以获得下一时间窗口内变量的权重。GRU 包含重置门 r_t 和更新门 z_t ,用于保留相关信息进行传递。基于注意力机制的 GRU 公式表达如下:

$$r_t = \sigma([W_r h_{t-1}; \tilde{h}_t^i]) \quad (13)$$

$$z_t = \sigma([W_z h_{t-1}; \tilde{h}_t^i]) \quad (14)$$

$$\tilde{g}_t = \tanh(W[r_t \odot h_{t-1}; \tilde{h}_t^i]) \quad (15)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{g}_t \quad (16)$$

其中, $[\cdot]$ 表示拼接操作, h_{t-1} 是上一时刻的隐藏状态, h_t 是传递到下一时刻的隐藏状态, $\tilde{g}_t$ 是候选隐藏状态, $\tilde{h}_t^i$ 是当前时刻的输入, $\sigma(\cdot)$ 是 sigmoid 函数, $\odot$ 是元素乘法, W_r 、 W_z 和 W_h 是可学习的参数。

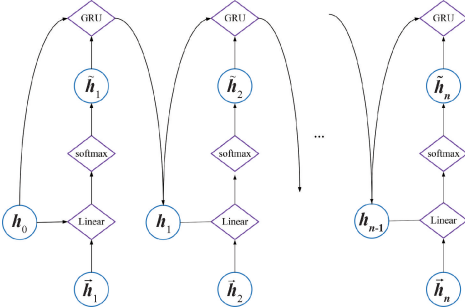


图 5 基于注意力机制 GRU 模块的结构

Fig. 5 Structure of the GRU module based on the attention mechanism

更具体地说,在每个 GRU 单元中,首先将 h_{t-1} 和 h_t^M 拼接起来,并输入线性层以获得 $\hat{w}^M$ 。然后经过 softmax 层转换,得到所有变量的最终权重 w^M 。公式表达如下:

$$\hat{w}_i = \tanh(W[h_{t-1}; \tilde{h}_t^i] + b) \quad (17)$$

$$w_i = \frac{\exp(\hat{w}_i)}{\sum_{j=1}^m \exp(\hat{w}_j)} \quad (18)$$

$$\tilde{h}_t^i = \tilde{h}_t^i \cdot w_i \quad (19)$$

其中, $\tilde{h}_t^i$ 是时间窗口 w_t 中第 i 个变量的向量表示, $\tilde{h}_t^i$

是基于注意力机制 GRU 模块的输出,然后将 h_{t-1} 和 $\tilde{h}_t^i$ 馈送到 GRU 单元来获取下一时间窗口的隐藏状态 h_t 。

由于 GRU 单元的输入是连续时间窗口的向量表示,因此基于注意力机制的 GRU 可以捕获时间依赖,并计算连续时间窗口内变量的权重,从而提取到不同时间段内每个变量的重要性。

1.7 联合优化模型

F-GAT 的输出是形状 $m \times k$ 的矩阵, T-GAT 的输出是 $k \times m$ 的矩阵。为了方便拼接,将基于注意力机制 GRU 的输出转化为 $k \times m$ 的形式,拼接后得到 $k \times 2m$ 的矩阵,其中每行表示每个时间戳的 $2m$ 维特征向量,融了时空相关性。

正如前面介绍到,基于预测和重构可以优势互补。使用基于多层感知机 (multilayer perceptron, MLP) 的预测模块和基于 GRU 自编码器的重建模块共同生成异常得分,损失函数定义为:

$$Loss = Loss_{pre} + Loss_{rec} \quad (20)$$

其中, $Loss_{pre}$ 表示预测模块的损失函数, $Loss_{rec}$ 表示重构模块的损失函数。

1) 预测模块

基于预测模块是预测下一个时间戳的观测值,将 GRU 的输出通过多层感知机 (MLP) 实现预测,损失函数为均方根误差 (root mean square error, RMSE):

$$Loss_{pre} = \sqrt{\sum_{i=1}^m (x_{t+1,i} - \hat{x}_{t+1,i})^2} \quad (21)$$

其中, $x_{t+1,i}$ 表示传感器 i 在 $t+1$ 时间戳的观测值, $\hat{x}_{t+1,i}$ 表示传感器 i 在 $t+1$ 时间戳的预测值。计算预测值与观测值的均方根误差作为异常评分。

2) 重构模块

基于重构模块学习整个时间序列的潜在表示,重构数据的正常模式。为了防止过拟合,对模型进行了简化。把 GRU 的输出作为潜在表示,再次利用 GRU 进行解码重构,其 RMSE 表达式如下:

$$Loss_{rec} = \sqrt{\sum (w_t - \hat{w}_t)^2} \quad (22)$$

其中, $\hat{w}_t$ 表示时间窗口 w_t 的重构值。

3) 异常检测

联合优化后,每个时间戳生成两个结果。通过预测的模型得到预测值 $\hat{x}_i, i \in R^m$, 通过重构的模型得到重构值 $\hat{x}_j, j \in R^m$ 。其中 $\hat{x}_i$ 表示第 i 个传感器时间序列的预测值。最终的异常评分平衡了两个模块的权重,每个时间戳的异常分数是所有传感器异常分数的平均值。异常分数可以通过以下方式计算:

$$score = \frac{1}{m} \left(\sqrt{\sum_{i=1}^m (x_{t+1,i} - \hat{x}_{t+1,i})^2} + \gamma \sqrt{\sum (w_t - \hat{w}_t)^2} \right) \quad (23)$$

其中, $\sqrt{\sum_{i=1}^m (x_{t+1,i} - \hat{x}_{t+1,i})^2}$ 为预测模型的异常分数, $\sqrt{\sum (w_t - \hat{w}_t)^2}$ 为重构模型的异常分数, γ 是一个平衡预测和重构的超参数, 由验证集选择。

最优全局阈值和之前的工作类似, 本文也采用 best-F1^[15]。当某个时间戳的异常得分超过阈值时, 则该时间戳被标记为“异常”, 否则就为“正常”。

2 实验结果与分析

为了验证模型的有效性, 本文进行了广泛的实验。首先介绍了 3 个常用的公共数据集, 并在这些数据集上进行评估。与基准模型进行对比, 证明了该模型优于当前的异常检测方法。然后使用消融实验对本文所提出的模块进行验证。

2.1 数据集

使用了 3 个真实世界的数据集来验证 STGAD 所提出模型的性能, 即火星科学实验室漫游车 (MSL)^[6]、土壤水分主动式被动 (SMAP)^[6] 和服务机器人数据 (SMD)^[16]。

火星科学实验室 (MSL) 探测车和土壤水分主动被动 (SMAP) 卫星是 NASA 公开发布的两个真实世界的数据集。每个数据集都包含一个训练集和一个测试集, 测试集中的异常由专家标记。MSL 有 27 个实体, 每个实体都监测了 55 个指标 (变量)。SMAP 有 55 个实体, 每个实体都监测了 25 个指标 (变量)。

SMD 是一家大型互联网公司收集的真实世界服务器机器数据集。SMD 包含服务器的时间序列数据, 该数据集来自 28 台机器, 每个服务器有 38 个多变量。它被划分为大小相等的训练集和测试集。

数据统计表如表 1 所示, 其中包括了训练集测试集的大小, 实体和特征数目, 以及测试集中的异常占比。

表 1 数据统计表

Table 1 The statistical table of datasets

数据	训练集	测试集	# 实体	# 特征	异常率/%
MSL	58 317	73 729	27	55	10.72
SMAP	135 183	427 617	55	25	13.13
SMD	708 405	708 420	28	38	4.16

在进行实验前, 对数据集进行预处理, 以提高模型的鲁棒性。对训练集和测试集都进行了归一化处理, 以训练集为例:

$$\tilde{x} = \frac{x - \min(X_{train})}{\max(X_{train}) - \min(X_{train})} \quad (24)$$

其中, $\min(X_{train})$ 和 $\max(X_{train})$ 分别表示训练集的最小值和最大值。

2.2 评价指标

本文使用精确率 (Precision)、召回率 (Recall)、F1 分数

(F1-score) 作为性能指标, 来评估该模型和基线的性能。计算方式为:

$$Precision = \frac{TP}{TP + FP} \quad (25)$$

$$Recall = \frac{TP}{TP + FN} \quad (26)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (27)$$

其中, TP 表示真阳性, FP 表示假阳性, FN 表示假阴性。由于异常通常持续时间比较短, 采用点调整法^[17]来检测模型的性能。该方法提出, 即使有一个观测值被检测为异常, 那么这一段的所有其他观测值也被检测为异常, 并且认为分类是正确的。

2.3 基线方法

为了评估该方法的性能, 将 STGAD 与以下十种流行的无监督异常检测方法进行了比较。

1) LSTM-NDT^[6] 使用了 LSTM 预测航天器的遥测数据, 通过计算误差检测异常, 同时引入了一种无监督的非参数异常阈值策略。

2) LSTM-VAE^[18] 提出了一种组合模型, 它将 VAE 学习到的潜在变量送入到 LSTM 捕获时间依赖, 最后利用重构概率确定异常。

3) AE^[7] 包括一个编码器和一个解码器, 并使用重构误差来检测异常。编码器将输入数据压缩为一个隐藏向量, 解码器使用该向量重构输入数据。

4) OmniAnomaly^[16] 将 GRU 和 VAE 联合起来, 利用平面归一化流和随机变量连接来提升模型效果。

5) USAD^[19] 包括一个编码器和两个解码器, 通过对抗训练拉大正常和异常数据间的差距, 从而能够检测异常。

6) GDN^[20] 利用基于图的嵌入和注意机制, 通过学习数据内部复杂关系和偏差, 预测未来数据检测异常。

7) MTAD-GAT^[12] 是一个比较全面的框架, 它包含一个时间图和一个空间图, 利用并行的 GAT 来学习复杂的时空关系, 同时优化预测和重构两个模块进行异常检测。

8) MST-GAT^[21] 借鉴了 GDN 的构图方式, 使用 3 个注意力模块提取模态间和模态内的相关性, 从而深入探索时间序列的特征。

9) GTA^[22] 将 Transformer 与图学习相结合, 捕捉数据中的动态关系和时间依赖性, 并利用 Transformer 对时间序列的全局时间依赖进行建模。

10) DuoGAT^[23] 面向时间建模, 利用图注意力网络对时间变化趋势进行学习, 注重捕捉时间的变化趋势。

2.4 实验环境

所有实验使用 PyTorch1.7.1 版本与 CUDA11.0 和 PyTorch 几何库 1.5.0 版本, 并在装有第 11 代 Intel CPU、3.50 GHz i9-11900KF 和 NVIDIA GeForce RTX 3090 GPU 的服务器上进行了实验。使用了多达 50 个 epoch 进行训练,

设置早停法为 10,采用学习率为 1×10^{-3} 的 Adam 优化器。批次大小为 256,时间窗口大小为 {25,100,100},稀疏图的 K 值为 {20,15,15},GRU 层和全连接层的维度都设置为 150。

2.5 性能比较

以 SMAP 数据集为例,图 6 展示了模型准确检测异常的能力。其中蓝色背景表示预测异常数,红色背景表示真实异常,两者混合的紫色背景表示准确检测。

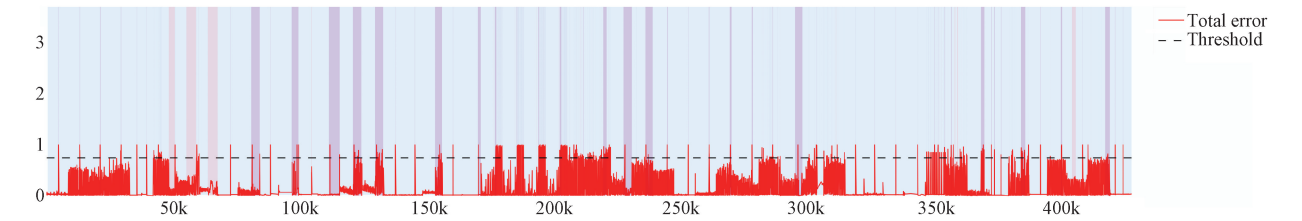


图 6 在 SMAP 测试集上的实验结果

Fig. 6 Experimental results on the SMAP test set

表 2 给出了该模型和基准模型在 3 个数据集上的性能指标,STGAD 在每种数据集上都表现出了最先进的性能。

与第二高的 DuoGAT 相比,F1 分数在 3 个数据集上分别提高了 1.5%、1.3%和 1.8%。

表 2 所提出的模型和基准模型实验结果

Table 2 Experimental results of the proposed and benchmark models

方法	MSL			SMAP			SMD		
	精确率	召回率	F1 分数	精确率	召回率	F1 分数	精确率	召回率	F1 分数
LSTM-VAE	0.525 7	0.954 6	0.678 0	0.855 1	0.636 6	0.729 8	0.887 3	0.511 1	0.648 6
LSTM-NDT	0.593 4	0.537 4	0.564 0	0.896 5	0.884 6	0.890 5	0.568 4	0.643 8	0.603 7
AE	0.716 6	0.500 8	0.589 6	0.721 6	0.799 5	0.758 6	0.777 8	0.510 9	0.616 7
OmniAnomaly	0.886 7	0.911 7	0.898 9	0.741 6	0.977 6	0.843 4	0.833 4	0.944 9	0.885 7
USAD	0.930 8	0.891 7	0.910 8	0.909 6	0.852 9	0.880 3	0.998 9	0.802 6	0.890 0
GDN	0.913 5	0.861 2	0.886 6	0.893 2	0.887 2	0.890 2	0.717 0	0.997 4	0.834 2
MTAD-GAT	0.875 4	0.944 0	0.908 4	0.890 6	0.912 3	0.901 3	0.940 9	0.852 4	0.894 5
MST-GAT	0.950 6	0.891 0	0.919 8	0.912 6	0.898 3	0.905 4	0.990 8	0.843 9	0.911 5
GTA	0.910 4	0.911 7	0.911 1	0.891 1	0.917 6	0.904 1	0.834 6	0.999 9	0.909 8
DuoGAT	0.927 1	0.953 8	0.940 3	0.863 4	1.000 0	0.926 7	0.853 8	0.989 1	0.916 5
STGAD	0.955 0	0.954 2	0.954 6	0.942 8	0.934 8	0.938 8	0.883 2	0.989 2	0.933 2

通过观察可以发现,基于预测的方法(LSTM-NDT、LSTM-VAE)可以捕获时间上下文的依赖关系,然而在存在噪声的情况下鲁棒性较差。基于重构的方法(AE、OmniAnomaly、USAD)可以捕捉数据的隐藏信息来重构数据,但是容易出现过拟合异常,从而导致性能低下。STGAD 联合优化预测和重构模块,将两者的优势结合,起到了不错的效果。在基于图的方法(GDN、MTAD-GAT、MST-GAT、GTA、DuoGAT)中,主要使用图神经网络进行特征关联,GDN 只建模了空间图,没有捕捉时间特征。MST-GAT 和 GTA 同样都只基于传感器的图结构进行学习。DuoGAT 对时间特性进行建模,考虑了时间变化趋势,但缺少了对传感器特征关系的学习。MTAD-GAT 全面地使用了面向特征和时间的图结构,然而这两个是无向无权的完整图,不能反应特征和时间的特性。为了剔除图结构的信息冗余,STGAD 对时空特性充分考虑,设计了两种图结构,更好地捕获时空依赖关系。此外,它们都忽略

了不同变量在检测异常时重要性是不同的,STGAD 关注了这一点,提高了检测精度,降低了误报和漏报的发生率。

2.6 消融实验

为了证明各个模块在模型中的有效性,在所有数据集上进行了消融实验。主要做的改进包括:除去多尺度门控卷积(记为 w/o GRes2Net),将稀疏空间图换成无向无权的完整图结构(记为 w/o A_{ij}),将时间导向图换成无向无权的完整图结构(记为 w/o B_{ij}^w),使用普通 GRU 代替基于注意力机制的 GRU(记为 w/o A_u)。结果如图 7 所示。观察结果,可以得出以下结论:

当移除 GRes2Net 模块时,性能出现了明显的下降。它说明了提取深层次特征的重要性,而且 GRes2Net 也能够强有力地捕获细粒度的局部特征。

同时也证明了设计的空间和时间图结构的有效性。仅对相关的传感器进行建模,去除了冗余,对于模型能够准确识别这种关系有很大帮助。采用有向且加权的时间

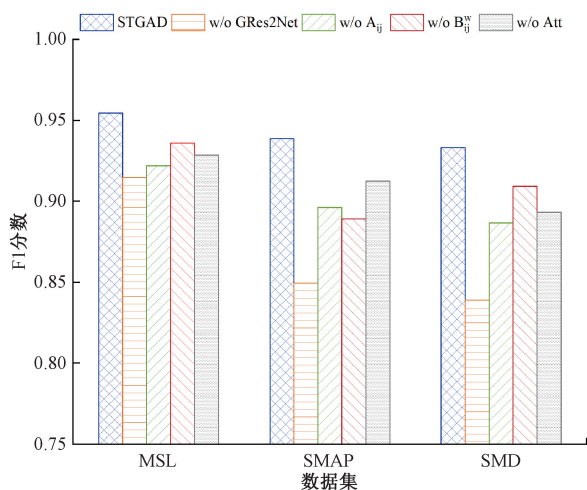


图 7 变体在 3 个数据集上的性能比较

Fig. 7 Performance comparison of variants on 3 datasets

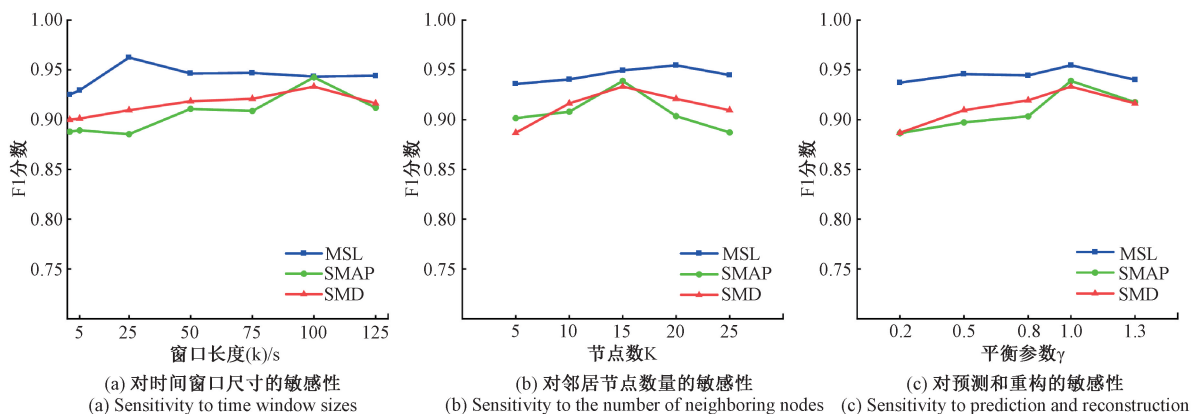


图 8 STGAD 参数在 3 个数据集上的实验结果

Fig. 8 Experimental results of STGAD parameters on three datasets

包含太多复杂的信息,变量之间的相互关系变得复杂,从而无法正确建模。

第 2 个参数是 STGAD 如何响应不同的稀疏参数 K 。 K 的值决定了有多少个与当前变量相似的变量通过连接一条边被选择为邻居节点。图 8(b) 给出了不同 K 选择下 STGAD 的性能。可以观察到, K 对模型的影响在时间窗长度 k 下具有类似的观察结果。当 K 的值达到 $\{20, 15, 15\}$ 时,模型性能达到最佳。只有选择合适的 K ,才能达到获取不同变量之间潜在联系的目的。如果 K 的值太小,一些变量之间的联系可能会被忽略,而当 K 的值太大时,变量之间建立更多的联系。但变量之间的关系强度会因为一些噪声而减弱,这并不能反映变量之间的真实联系。

最后,评估了超参数 γ 对 STGAD 在 3 个数据集上检测性能的影响。用 γ 来平衡预测和重构模块的重要性。实验结果如图 8(c) 所示。可以看出,随着超参数 γ 的增大,STGAD 的 F1 分数先增大到最大值,然后减小。这是直观的,因为在消融研究中证实,预测和重建模块对于精确检测是必不可少的。3 个数据集同时在 γ 设为 1 时, F1

图结构也很好地反映了时间特性。把这两种图结构替换为原来的无权无向完整的图结构,都有不同程度的性能下降。最后去除注意力机制,性能也在下降。这表明基于注意力机制的 GRU 在学习变量在每个时间段的重要性发挥了重要作用。

2.7 参数敏感性

本节对 STGAD 参数的敏感性进行研究,包括时间窗口长度 k , 节点邻居数 K 和平衡参数 γ 。所有实验均使用 3 个数据集进行,实验结果如图 8 所示。

图 8(a) 显示了通过改变时间窗长度从 1~125 得到的 STGAD 在 F1 分数方面的结果。可以看出,随着时间窗长度的增加,STGAD 的检测效果先增大后减小。

当 $k = \{25, 100, 100\}$ 时,STGAD 达到最佳性能。这是因为当时间窗口太小时,由于窗口内的数据有限,无法很好地捕捉变量之间的相关性。但当时间窗口太大时,它会

分数达到最大值,即 STGAD 的性能达到最佳。这也表明所提出的预测和重建模块对整体性能贡献很大。

3 结 论

针对现有的时间序列异常检测方法没有充分考虑时空特性,对时空图建模存在冗余,导致计算复杂度偏高,以及忽略了变量在不同时间段的重要性。提出了一种新的用于 MTS 异常检测的无监督方法 STGAD。STGAD 首先将门控机制引入多尺度卷积网络中,提取深层次的特征同时融合更相关的特征。然后设计了时间图 and 空间图,更好地建模时空相关性,提高了模型计算速度。通过 GTA 和 GRU 进一步学习时空依赖关系关系。为了进一步改进,提出了一种基于注意力机制的 GRU 模块,学习每个时间窗口中变量的重要性。最后下游任务由基于预测和重构模块联合进行异常检测。在 3 个公共数据集 (MSL、SMAP、SMD) 上与基准模型进行对比实验,在 F1 分数性能指标上分别有 1.5%、1.3% 和 1.8% 的提高,优于其他基准模型。证明了它可以应用于复杂系统,如网络物理系统

和污水处理等。在未来的工作中,将会进一步优化该模型的性能。引入一个轻量化 Transformer 模块,进行更高效率的异常检测。

参考文献

- [1] LI ZH H, ZHAO Y J, HAN J Q, et al. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding[C]. 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021: 3220-3230.
- [2] MOHAMMADI M, AL-FUQAHA A, SOROUR S, et al. Deep learning for IoT big data and streaming analytics: A survey [J]. IEEE Communications Surveys & Tutorials, 2018, 20(4): 2923-2960.
- [3] 刘义艳,郝婷楠,张伟.融合 RMT 特征值的电网异常状态检测[J].电子测量与仪器学报,2023,37(12): 242-252.
- LIU Y Y, HAO T N, ZHANG W. Abnormal state detection of power system based on RMT eigenvalue fusion[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(12):242-252.
- [4] ZHENG X, CAI ZH P. Privacy-preserved data sharing towards multiple parties in industrial IoTs[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(5): 968-979.
- [5] 郝椿淋,张剑.基于自注意力机制 TCN-BiGRU 的交通流预测[J].电子测量技术,2024,47(8):61-68.
- HAO CH L, ZHANG J. Traffic flow prediction based on self-attention mechanism TCN-BiGRU [J]. Electronic Measurement Technology, 2024, 47(8): 61-68.
- [6] HUNDMAN K, CONSTANTINOU V, LAPORTE C, et al. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding [C]. 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 387-395.
- [7] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]. International Conference on Learning Representations, 2018.
- [8] LI D, CHEN D CH, JIN B H, et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks[C]. International Conference on Artificial Neural Networks. Cham: Springer International Publishing, 2019: 703-716.
- [9] KIM H, LEE B S, SHIN W Y, et al. Graph anomaly detection with graph neural networks: Current status and challenges [J]. IEEE Access, 2022, 10: 111820-111829.
- [10] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. ArXiv preprint arXiv:1609.02907, 2016.
- [11] VELICKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. ArXiv preprint arXiv:1710.10903, 2017.
- [12] ZHAO H, WANG Y J, DUAN J Y, et al. Multivariate time-series anomaly detection via graph attention network [C]. 2020 IEEE International Conference on Data Mining(ICDM), 2020: 841-850.
- [13] GAO SH H, CHENG M M, ZHAO K, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 652-662.
- [14] 刘少卿,李帅,苗建国,等.基于 TCN-BiGRU 的锂离子电池健康状态评估[J].电子测量技术,2023,46(23): 68-76.
- LIU SH Q, L SH, MIAO J G, et al. Lithium-ion battery state of health estimation based on TCN-BiGRU [J]. Electronic Measurement Technology, 2023, 46(23):68-76.
- [15] GOH J, ADEPU S, TAN M, et al. Anomaly detection in cyber physical systems using recurrent neural networks[C]. 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), 2017: 140-145.
- [16] SU Y, ZHAO Y J, NIU CH H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 2828-2837.
- [17] XU H W, CHEN W X, ZHAO N W, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications[C]. 2018 World Wide Web Conference, 2018: 187-196.
- [18] PARK D, HOSHI Y, KEMP C C. A multimodal

anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1544-1551.

[19] AUDIBERT J, MICHIARDI P, GUYARD F, et al. Usad: Unsupervised anomaly detection on multivariate time series [C]. 26th ACM SIGKDD International Conference on Nknowledge Discovery & Data Mining, 2020: 3395-3404.

[20] DENG A, HOOI B. Graph neural network-based anomaly detection in multivariate time series [C]. AAAI Conference on Artificial Intelligence, 2021: 4027-4035.

[21] DING CH Y, SUN SH L, ZHAO J. MST-GAT: A multimodal spatial-temporal graph attention network for time series anomaly detection [J]. Information Fusion, 2023, 89: 527-536.

[22] CHEN Z K, CHEN D SH, ZHANG X, et al. Learning graph structures with transformer for multivariate time-series anomaly detection in IoT[J]. IEEE Internet of Things Journal, 2021, 9 (12): 9179-9189.

[23] LEE J, PARK B, CHAE D K. DuoGAT: Dual time-oriented graph attention networks for accurate, efficient and explainable anomaly detection on time-series [C]. 32nd ACM International Conference on Information and Knowledge Management, 2023: 1188-1197.

作者简介

杨晨龙, 硕士研究生, 主要研究方向为数据异常检测。
E-mail: yangchenlong@stu.ncst.edu.cn

孙晔(通信作者), 博士研究生, 研究员, 主要研究方向为大数据分析。
E-mail: sunyecaict@126.com

刘晓悦, 博士, 教授, 主要研究方向为信息处理与网络控制。
E-mail: 1113610654@qq.com