

基于随机 Transformer 的多维时间序列异常检测模型

霍纬纲, 梁锐, 李永华

(中国民航大学计算机科学与技术学院, 天津 300300)

摘要: 针对已有基于变分自编码器 (VAE) 的多维时间序列 (MTS) 异常检测模型无法在隐空间中传播随机变量间的长时依赖性问题, 提出了一种融合 Transformer 编码器和 VAE 的随机 Transformer MTS 异常检测模型 (ST-MTS-AD)。在 ST-MTS-AD 的推断网络中, Transformer 编码器产生的当前时刻 MTS 长时依赖特征和上一时刻随机变量的采样值被输入多层感知器, 由此生成当前时刻随机变量的近似后验分布, 实现随机变量间的时序依赖。采用门控转换函数 (GTF) 生成随机变量的先验分布, ST-MTS-AD 的生成网络由多层感知器重构 MTS 各时刻取值分布, 该多层感知器的输入为推断网络生成的 MTS 的长时依赖特征和随机变量近似后验采样值。ST-MTS-AD 基于变分推断技术学习正常 MTS 样本集分布, 由重构概率对数似然确定 MTS 异常片段。4 个公开数据集上的实验表明, ST-MTS-AD 模型比典型相关基线模型的 F1 分数有明显提升。

关键词: 随机 Transformer; 变分自编码器; 多维时间序列; 异常检测

中图分类号: TP391

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023042

Anomaly detection model for multivariate time series based on stochastic Transformer

HUO Weigang, LIANG Rui, LI Yonghua

School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China

Abstract: Aiming at the problem that the existing multivariate time series anomaly detection models based on variational autoencoders could not propagate long-term temporal dependencies between stochastic variables in latent space, the stochastic Transformer for MTS anomaly detection (ST-MTS-AD) model which combined Transformer encoder with VAE was proposed. In the inference network of the ST-MTS-AD, the MTS long-term temporal dependent features generated by Transformer encoder and the sampled values of the stochastic variables at the previous moment were inputted into the multilayer perceptron, the approximate posterior distribution of the stochastic variables at the current moment was generated by the multilayer perceptron, and the temporal dependencies between stochastic variables were realized. The gated transition function (GTF) was used to generate the prior distribution of stochastic variables. The generation network of the ST-MTS-AD reconstructed the distribution of the MTS values at each moment by the multilayer perceptron whose input was the MTS long-term temporal dependent features generated by the inference network and the approximate posterior sampling values of stochastic variables. The distribution of normal MTS dataset was learned by the variational inference technology, and the abnormal MTS segment was determined by the log-likelihood of the reconstruction probability. Experiments on four public datasets show that the ST-MTS-AD model significantly improves the F1 score over the typical baseline models.

Keywords: stochastic Transformer, variational autoencoder, multivariate time series, anomaly detection

收稿日期: 2022-08-30; 修回日期: 2022-11-21

通信作者: 李永华, yh-li@cauc.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62173331); 中央高校基本科研业务费专项资金资助项目 (No.3122019190)

Foundation Items: The National Natural Science Foundation of China (No.62173331), The Fundamental Research Funds for the Central Universities (No.3122019190)

0 引言

得益于传感器及监控技术的进步, 现代工业设备的生产运行状态和运行环境能够被实时记录和感知, 积累并产生了海量多维时间序列 (MTS, multivariate time series) 数据^[1]。例如, 云服务器的 CPU 利用率、内存占用率、网络流量等监测数据, 航空器的速度、高度、俯仰角度等传感器数据, 反映航天器运行状态的遥测信号, 工厂生产机器的温度、转速等监测数据。这些 MTS 数据的异常取值片段往往意味着设备故障、操作人员失误等特殊情况的发 生, 也可能表示生产环境中存在隐性安全隐患。这些异常如不能被有效识别, 很可能造成经济损失。工业领域 MTS 数据有如下特点: 体量大、连续采样、价值密度低、动态性强; MTS 各维度之间具有复杂的时空耦合关系; 工业设备通过其软件控制逻辑往往与其运行环境、操作人员及相关设备系统有复杂交互, 反映其运行状态的 MTS 取值具有随机性^[2]。另外, 由于工业领域 MTS 异常样本相对较少、标注异常代价较高, 研究者大多关注基于无监督的 MTS 异常检测方法。传统的主成分分析、高斯混合模型、一类支持向量机等机器学习方法无法很好地建模工业领域 MTS 数据的复杂性质。由于深度学习强大的数据表征能力, 近年来, 基于深度学习的无监督 MTS 异常检测研究得到了广泛关注。MTS 异常检测包括序列级和瞬时级异常发现。序列级异常^[3-5]是指 MTS 样本整个序列或其子序列区别于多数样本。瞬时级异常是指在某个时间点或短时间窗内的异常。本文模型属于瞬时级异常检测模型。

围绕如何建模工业领域 MTS 时序依赖性和随机性, 研究者提出了一些深度学习模型^[2,6-10]。这些模型的训练目标均为学习正常 MTS 样本集的时序分布。其中, 大部分方法均结合循环神经网络 (RNN, recurrent neural network) 和变分自编码器 (VAE, variational autoencoder) 建模 MTS 数据的时序依赖性和随机性。但已有模型存在如下问题。

1) 采用 RNN 的隐向量实现 VAE 隐空间中随机变量间的时序依赖性, 然而 RNN 难以捕获序列数据的长时依赖性, 这降低了学习序列数据分布的能力。

2) 计算随机变量的近似后验分布和先验分布的网络结构相同, 这使 2 种分布 KL 散度 (Kullback-Leibler divergence) 距离较小, 增加了模型的训练

难度。基于线性高斯状态空间模型 (LGSSM, linear Gaussian state space model) 计算随机变量的先验分布无法实现随机变量间的非线性转换。

3) 基于 RNN 的生成网络仅依赖于随机变量的采样值, 没有利用 RNN 推断网络的确定性隐向量。

针对以上问题, 本文提出一种面向工业领域 MTS 异常检测的随机 Transformer (ST-MTS-AD, stochastic Transformer for MTS anomaly detection) 模型。该模型由 Transformer 编码器输出的表示各时刻 MTS 长时依赖特征和上一时刻随机变量的采样值生成当前时刻随机变量的近似后验分布, 使 ST-MTS-AD 可借用 Transformer 编码器输出的长时依赖特征在隐空间中传播随机变量间的长时依赖性, 采用门控转换函数 (GTF, gated transition function) 生成随机变量的先验分布并实现随机变量间的非线性转换, 通过将 Transformer 编码器输出的长时依赖特征和随机变量采样值输入多层感知器 (MLP, multilayer perceptron) 重构 MTS 各时刻取值分布。在 4 个工业领域公开 MTS 数据集上的实验表明 ST-MTS-AD 具有较好的异常检测效果。

1 相关研究现状

多维时间序列异常检测是时间序列分析领域中的重要任务之一, 旨在寻找不符合规则的或出现偏差的序列片段^[11]。目前, 基于深度学习的 MTS 瞬时级异常检测方法如下。

1) 基于预测和重构的确定性方法。Hundman 等^[12]设计了基于长短时记忆 (LSTM, long short-term memory) 网络的航天器遥测信号异常检测方法, 通过预测误差确定异常, 其还提出了一种非参数动态异常检测阈值确定方法, 能在误报率和漏报率之间达到平衡。Zhang 等^[13]提出了面向多传感器 MTS 异常检测的深度卷积自编码记忆网络, 将 MTS 的时空嵌入表示和重构误差输入线性自回归模型和基于注意力机制的双向 LSTM 网络, 由重构损失和预测损失确定 MTS 异常。Malhotra 等^[14]提出了一种基于 LSTM 的自编码器模型, 旨在重建正常时间序列, 使用重构误差进行异常检测。Zhang 等^[15]提出了一种多尺度卷积循环自编码器 (MSCRED), 首先构建能表征 MTS 不同变量间相关性的多尺度签名矩阵; 然后采用卷积编码器对签名矩阵进行编码, 使用基于注意力的卷积 LSTM 捕获 MTS 时序依赖性; 最后基于卷积解码器重建签

名矩阵, 并利用签名矩阵的重构误差诊断异常。Audibert 等^[16]设计了包含一个编码器和 2 个解码器的自编码器网络结构, 采用对抗学习策略训练网络, 避免自编码器无法通过重构误差区分正常样本与异常样本的现象, 该方法没有使用循环神经网络从而获得了较快的训练效率, 但无法建模 MTS 的时序依赖关系。上述方法以预测或重构为训练目标刻画正常 MTS 样本的特征, 由重构误差和预测误差检测 MTS 异常, 无法建模 MTS 的随机性。

2) 基于重构的随机性方法。Zong 等^[6]提出一种用于无监督异常检测的深度自编码高斯混合模型 (DAGMM, deep autoencoding Gaussian mixture model), 通过自编码器获得训练样本的低维表示, 将低维表示和样本重构误差拼接形成的向量输入用来训练 GMM 参数的估计网络, 由 GMM 模型计算的样本能量值确定样本是否异常。Deng 等^[7]采用图卷积网络捕获 MTS 变量间交互特征, 由自注意力机制提取 MTS 长时依赖特征, 基于 VAE 的重构能力进行 MTS 异常检测。Park 等^[8]提出了基于 LSTM 和 VAE 的多模态 MTS 异常检测方法, 以 LSTM 网络结构作为 VAE 中的生成网络和推断网络表示 MTS 的随机性和时序依赖性。文献[7-8]通过自注意力机制或 RNN 捕获 MTS 的时序依赖性, 并且每个时间点通过重参数化采样得到的随机变量具有随机性, 但随机变量之间没有时序依赖性。

为此, 研究者提出了基于 VAE 的序列数据生成模型^[17-19], 这些方法均采用变分推断技术学习时序数据分布, 并且可以表示随机变量间的时序依赖性。Chung 等^[18]提出一种变分循环神经网络 (VRNN, variational recurrent neural network) 模型。VRNN 的推断网络由时序数据当前时刻输入值和上一时刻 RNN 隐向量生成当前时刻随机变量的近似后验分布。VRNN 通过 RNN 隐向量的迭代更新实现随机变量间的时序依赖性。VRNN 的生成网络由当前时刻随机变量采样值和上一时刻 RNN 隐向量生成时序数据取值分布。Dai 等^[9]提出了面向内容分发网络系统多元关键性能指标时序数据的异常检测方法 SDFVAE (static and dynamic factorized VAE), 将 VRNN 中的随机变量分解为动态和静态随机变量, 其中, 静态随机变量刻画内容分发网络系统性能指标取值的时序不变性。SDFVAE 由双向 LSTM 网络学习静态随机变量的近似后验分布, 动态随机变量的近似后验分布生成方式与 VRNN 相同。Li

等^[10]提出了基于 VRNN 的 MTS 异常检测方法, 在损失函数中引入使相邻时刻点时序取值生成分布相近的正则化项, 使模型具有更好的抗噪能力。Fraccaro 等^[19]提出了一种融合状态空间模型 (SSM, state space model) 和 RNN 的随机循环神经网络 (SRNN, stochastic recurrent neural network)。SRNN 模型通过依赖于 RNN 隐向量和 SSM 随机变量采样值的神经网络实现 SSM 随机状态之间的非线性转换, 使 SSM 可以利用 RNN 表示时序依赖关系的隐向量在隐空间中传播随机性。SRNN 与 VRNN 的区别在于 RNN 隐向量的更新不依赖于各时刻的随机变量, 能直接建模每个时刻随机变量间的时序依赖性, 实现了 RNN 确定性隐变量和 SSM 随机变量的分离。文献[19]表明 SRNN 能够比 VRNN 更好地学习时序数据的分布。Su 等^[2]提出基于 SRNN 的 MTS 异常检测模型, 该模型的推断网络结构与 SRNN 相同, 在此基础上采用平面归一化流技术学习各时刻非高斯分布的近似后验分布。在生成网络中, 该模型由依赖于随机变量采样值的 RNN 生成时序数据取值分布, 采用 LGSSM 计算随机变量的先验分布并实现随机变量间的时序依赖性。

2 ST-MTS-AD 模型设计

2.1 相关符号定义及 ST-MTS-AD 网络结构

MTS 数据集记为 $\mathbf{k} = [\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_N] \in \mathbb{R}^{M \times N}$, N 为 \mathbf{k} 的持续时间长度, 每个观测值 $\mathbf{k}_\tau \in \mathbb{R}^M$ 是在时间点 $\tau (\tau \leq N)$ 的 M 维向量。使用窗口大小为 w 、滑动步幅为 l 的滑动窗口对 \mathbf{k} 进行预处理, 每个滑动窗口为

$$\mathbf{x}_j = [\mathbf{k}_{\tau-w+1}, \dots, \mathbf{k}_{\tau-1}, \mathbf{k}_\tau] \in \mathbb{R}^{M \times w}$$

其中, $1 \leq j \leq \left\lfloor \frac{(N-w)}{l} \right\rfloor + 1$, $\lfloor \cdot \rfloor$ 表示向下取整。

将 \mathbf{x}_j 称为观测变量, T 个 \mathbf{x}_j 拼接形成观测序列, 由此 MTS 数据集 \mathbf{k} 可形成 $A = \left\lfloor \frac{N-w+l}{lT} \right\rfloor$ 个观测序列, 每个观测序列记为 $\mathbf{x}_{1:T}^i = [\mathbf{x}_{1:T}^i, \dots, \mathbf{x}_t^i, \dots, \mathbf{x}_T^i] \in \mathbb{R}^{P \times T}$, 其中, $1 \leq i \leq A$, $P=Mw$, \mathbf{x}_t^i 表示第 i 个观测序列中的第 $t (1 \leq t \leq T)$ 个观测变量。所有观测序列构成的集合表示为 $D(\mathbf{x}_{1:T}) = \{\mathbf{x}_{1:T}^i\}_{i=1}^A$, 这些序列的数据分布表示为 $p_D(\mathbf{x}_{1:T})$ 。异常检测模型的训练样本简写为 $\mathbf{x}_{1:T}$, 异常检测目标为判断观测变量 \mathbf{x}_t 是否异常。

本文设计的 ST-MTS-AD 模型基于 VAE 的变分

推断技术学习 MTS 时序分布, 其主要优势在于使用 Transformer 编码器中的多头自注意力机制生成 MTS 各个时刻点的随机分布, 实现了 VAE 隐空间中随机变量间的长时依赖关联。而且 ST-MTS-AD 的生成网络为 MLP, 降低了模型的复杂性。ST-MTS-AD 模型结构如图 1 所示。在推断网络中, Transformer 编码器将观测序列 $\mathbf{x}_{1:T} \in \mathbb{R}^{P \times T}$ 映射为特征 $\mathbf{e}_{1:T} = [\mathbf{e}_1, \dots, \mathbf{e}_t, \dots, \mathbf{e}_T] \in \mathbb{R}^{d \times T}$, d 为 Transformer 编码器输出维度, $\mathbf{e}_{1:T}$ 捕获 $\mathbf{x}_{1:T}$ 中观测变量 x_i 间的长时依赖性, $\mathbf{e}_t (1 \leq t \leq T)$ 表示 $\mathbf{e}_{1:T}$ 在 t 时刻的取值向量。将 $t-1$ 时刻随机变量 z_{t-1} 的采样值和 \mathbf{e}_t 串接后输入 MLP 生成 t 时刻随机变量 z_t 的近似后验分布, 实现随机变量 z_{t-1} 和 z_t 的非线性转换和时序依赖关联。记 T 个随机变量 z_1, \dots, z_T 为 $\mathbf{z}_{1:T}$ 。在生成网络中, t 时刻随机变量 z_t 采样值和取值向量 \mathbf{e}_t 输入 MLP 生成 \mathbf{x}'_t 的取值分布。 T 个时刻观测变量 x_i 的重构值 $\mathbf{x}'_1, \dots, \mathbf{x}'_T$ 表示为 $\mathbf{x}'_{1:T}$ 。 $t-1$ 时刻随机变量 z_{t-1} 的采样值输入 GTF 生成 t 时刻随机变量 z_t 的先验分布, z_0 为随机初始化的向量。ST-MTS-AD 通过最大化证据下界优化推断网络和生成网络参数, 由各时刻 \mathbf{x}'_t 的重构概率似然确定 $\mathbf{x}_{1:T}$ 异常片段。

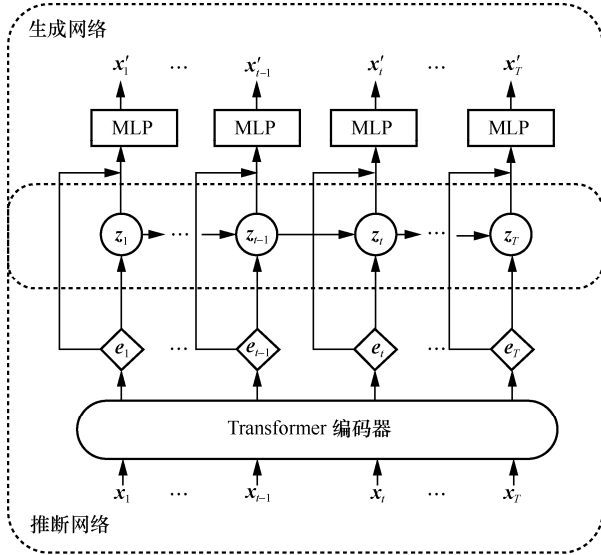


图 1 ST-MTS-AD 模型结构

2.2 ST-MTS-AD 推断网络设计

Transformer 网络中的自注意力机制容易捕获序列数据中长时依赖特征, 文献 [20] 基于 Transformer 编码器设计了 MTS 的高质量表示学习模型。ST-MTS-AD 模型将 Transformer 编码器作为 VAE 的推断网络, 捕获观测序列 $\mathbf{x}_{1:T}$ 中各观测变量

x_i 间的长时依赖性。为了实现随机变量 z_{t-1} 和 z_t 的非线性转换和时序依赖关联, 将 $t-1$ 时刻随机变量 z_{t-1} 的采样值和 Transformer 编码器在 t 时刻的输出 \mathbf{e}_t 串接后生成 t 时刻随机变量 z_t 。推断网络的目标是学习随机变量 $\mathbf{z}_{1:T}$ 的真实后验分布, ST-MTS-AD 通过推断网络得到 $\mathbf{z}_{1:T}$ 的近似后验分布 $q_\phi(\mathbf{z}_{1:T} | \mathbf{x}_{1:T})$, 如式(1)所示。

$$q_\phi(\mathbf{z}_{1:T} | \mathbf{x}_{1:T}) = \prod_{t=1}^T q_\phi(z_t | z_{t-1}, \mathbf{x}_{1:T}) \quad (1)$$

ST-MTS-AD 模型推断网络结构如图 2 所示, 即式(1)中 t 时刻近似后验分布 $q_\phi(z_t | z_{t-1}, \mathbf{x}_{1:T})$ 网络结构。

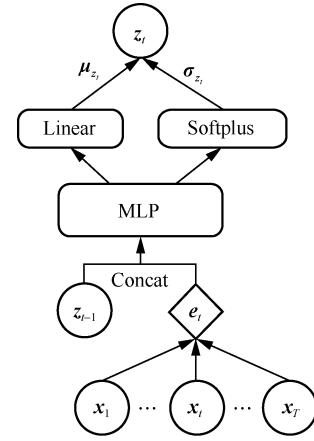


图 2 ST-MTS-AD 模型推断网络结构

$\mathbf{e}_{1:T}$ 计算过程如下。按式(2)对 $\mathbf{x}_{1:T}$ 进行位置编码, 编码结果记为 $\mathbf{x}_{1:T}^p \in \mathbb{R}^{d \times T}$ 。

$$\mathbf{x}_{1:T}^p = (\mathbf{W}_x \mathbf{x}_{1:T} + \mathbf{b}) + \mathbf{w}_p \quad (2)$$

其中, $\mathbf{W}_x \in \mathbb{R}^{d \times P}$ 为网络参数; $\mathbf{b} \in \mathbb{R}^{1 \times T}$ 为偏置项; $\mathbf{w}_p \in \mathbb{R}^{d \times T}$ 为位置编码矩阵, 由正弦函数在偶数位置计算得到的位置信息和余弦函数在奇数位置计算获得的位置信息结合形成, 具体形式如下

$$\mathbf{w}_p = \begin{bmatrix} \sin(w_1) & \dots & \sin(w_1 t) & \dots & \sin(w_1 T) \\ \cos(w_1) & \dots & \cos(w_1 t) & \dots & \cos(w_1 T) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sin(w_k) & \dots & \sin(w_k t) & \dots & \sin(w_k T) \\ \cos(w_k) & \dots & \cos(w_k t) & \dots & \cos(w_k T) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sin\left(\frac{w_d}{2}\right) & \dots & \sin\left(\frac{w_d t}{2}\right) & \dots & \sin\left(\frac{w_d T}{2}\right) \\ \cos\left(\frac{w_d}{2}\right) & \dots & \cos\left(\frac{w_d t}{2}\right) & \dots & \cos\left(\frac{w_d T}{2}\right) \end{bmatrix} \quad (3)$$

其中, $w_k = \frac{1}{10\,000^{\frac{2k}{d}}}$, $1 \leq k \leq \frac{d}{2}$, $1 \leq t \leq T$ 。

由式(4)~式(10)所示的多头注意力机制计算 $\mathbf{x}_{1:T}$ 中观测变量间长时依赖特征 $\mathbf{e}_{1:T}$ 。

$$\mathbf{Q}_h = \mathbf{x}_{1:T}^p \mathbf{W}_h^Q \quad (4)$$

$$\mathbf{K}_h = \mathbf{x}_{1:T}^p \mathbf{W}_h^K \quad (5)$$

$$\mathbf{V}_h = \mathbf{x}_{1:T}^p \mathbf{W}_h^V \quad (6)$$

$$\mathbf{O}_h = \text{Attention}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \text{softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^T}{\sqrt{d_K}}\right) \mathbf{V}_h \quad (7)$$

$$\mathbf{G}_{1:T} = \text{Concat}(\mathbf{O}_1, \dots, \mathbf{O}_H) \mathbf{W}^O \quad (8)$$

$$\mathbf{I}_{1:T} = \text{LN}(\mathbf{G}_{1:T} + \mathbf{x}_{1:T}^p) \quad (9)$$

$$\mathbf{e}_{1:T} = \text{LN}(\text{FFN}(\mathbf{I}_{1:T})) + \mathbf{I}_{1:T} \quad (10)$$

式(4)~式(6)通过网络参数矩阵 \mathbf{W}_h^Q 、 \mathbf{W}_h^K 和 \mathbf{W}_h^V 将 $\mathbf{x}_{1:T}^p$ 分别映射为查询矩阵 \mathbf{Q}_h 、键矩阵 \mathbf{K}_h 和值矩阵 \mathbf{V}_h 。多头自注意力机制的注意力头数记为 H 。 \mathbf{O}_h 为第 $h(h \in [1, H])$ 个子头的自注意力结果, d_K 是键矩阵的行数。Concat(\cdot) 为矩阵的拼接操作, \mathbf{W}^O 为网络参数。由式(9)和式(10)所示的残差网络结构计算得到特征 $\mathbf{e}_{1:T}$, 其中, LN(\cdot) 表示归一化层, FFN(\cdot) 表示前馈神经网络。

t 时刻推断网络生成的近似后验分布为

$$q_\phi(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{x}_{1:T}) = \mathcal{N}(\mathbf{z}_t | \psi_{\mu_{z_t}}(\mathbf{z}_{t-1}, \mathbf{e}_t), \text{diag}(\psi_{\sigma_{z_t}}(\mathbf{z}_{t-1}, \mathbf{e}_t))) \quad (11)$$

其中, \mathbf{z}_t 服从对角多元高斯分布; $\psi_{\mu_{z_t}}(\cdot)$ 表示图 2 中 MLP 和线性 (Linear) 层组成的神经网络, 该神经网络用于生成随机变量 \mathbf{z}_t 的均值。 $\psi_{\sigma_{z_t}}(\cdot)$ 表示图 2 中 MLP 和一个具有 Softplus 激活函数的全连接层组成的神经网络, 该神经网络生成随机变量 \mathbf{z}_t 的标准方差。神经网络 $\psi_{\mu_{z_t}}(\cdot)$ 和 $\psi_{\sigma_{z_t}}(\cdot)$ 的输入为 $t-1$ 时刻随机变量 \mathbf{z}_{t-1} 的采样值和 \mathbf{e}_t , 其中 \mathbf{z}_0 为随机初始化的向量。

2.3 ST-MTS-AD 生成网络设计

ST-MTS-AD 生成网络由 MLP 构成, 目的是重构观测序列 $\mathbf{x}'_{1:T}$ 。其与文献[2,9]不同之处在于, 由随机变量 \mathbf{z}_t 采样值和推断网络中 Transformer 编码器在 t 时刻的输出 \mathbf{e}_t 生成 t 时刻的观测变量 \mathbf{x}'_t , 而不仅依赖随机变量 \mathbf{z}_t 。这是由于 \mathbf{e}_t 包含了来自输入观测序列 $\mathbf{x}_{1:T}$ 的全局时序特征, 能够更好地重构观测序列。同时, MLP 网络结构简单, 降低了 ST-MTS-AD 网络的复杂性。ST-MTS-AD 生成网络

的联合概率分布 $p_\theta(\mathbf{x}'_{1:T}, \mathbf{z}_{1:T})$ 为

$$p_\theta(\mathbf{x}'_{1:T}, \mathbf{z}_{1:T}) = p_\theta(\mathbf{x}'_{1:T} | \mathbf{z}_{1:T}) p_\theta(\mathbf{z}_{1:T}) = \prod_{t=1}^T p_\theta(\mathbf{x}'_t | \mathbf{z}_t) p_\theta(\mathbf{z}_t | \mathbf{z}_{t-1}) \quad (12)$$

其中, $p_\theta(\mathbf{x}'_t | \mathbf{z}_t)$ 为 t 时刻观测变量 \mathbf{x}'_t 的生成分布, $p_\theta(\mathbf{z}_t | \mathbf{z}_{t-1})$ 为 t 时刻随机变量 \mathbf{z}_t 的先验分布。ST-MTS-AD 模型生成网络结构如图 3 所示, 即式(12)中 $p_\theta(\mathbf{x}'_t | \mathbf{z}_t)$ 分布的网络结构。

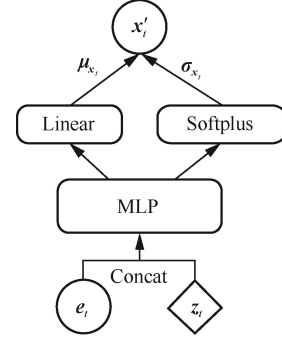


图 3 ST-MTS-AD 模型生成网络结构

t 时刻 $p_\theta(\mathbf{x}'_t | \mathbf{z}_t)$ 分布形式如下

$$p_\theta(\mathbf{x}'_t | \mathbf{z}_t) = \mathcal{N}(\mathbf{x}'_t | f_{\mu_x}(\mathbf{z}_t, \mathbf{e}_t), \text{diag}(f_{\sigma_x}(\mathbf{z}_t, \mathbf{e}_t))) \quad (13)$$

其中, \mathbf{x}'_t 服从对角多元高斯分布; $f_{\mu_x}(\cdot)$ 表示图 3 中 MLP 和 Linear 层组成的神经网络, 该网络生成随机变量 \mathbf{x}'_t 的均值; $f_{\sigma_x}(\cdot)$ 表示图 3 中 MLP 和具有 Softplus 激活函数的全连接层组成的神经网络, 该网络生成 \mathbf{x}'_t 的标准方差。神经网络 $f_{\mu_x}(\cdot)$ 和 $f_{\sigma_x}(\cdot)$ 的输入为随机变量 \mathbf{z}_t 采样值和 \mathbf{e}_t 。按式(14)所示的重采样过程得到 \mathbf{z}_t 采样值。

$$\mathbf{z}_t = \boldsymbol{\mu}_{z_t} + \boldsymbol{\xi} \odot \boldsymbol{\sigma}_{z_t} \quad (14)$$

其中, $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, \odot 表示向量元素乘积, $\boldsymbol{\mu}_{z_t}$ 和 $\boldsymbol{\sigma}_{z_t}$ 分别表示由图 2 推断网络生成的 \mathbf{z}_t 近似后验分布均值和标准方差。

采用如图 4 所示的 GTF^[17] 生成随机变量 \mathbf{z}_t 的先验分布 $p_\theta(\mathbf{z}_t | \mathbf{z}_{t-1})$, 并实现随机变量间的非线性转换, 如式(15)所示。

$$p_\theta(\mathbf{z}_t | \mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t | \mathbf{g}_{\mu_{z_t}}(\mathbf{z}_{t-1}), \text{diag}(\mathbf{g}_{\sigma_{z_t}}(\mathbf{z}_{t-1}))) \quad (15)$$

其中, 随机变量 \mathbf{z}_t 服从对角多元高斯分布, 其均值和标准方差分别由神经网络 $\mathbf{g}_{\mu_{z_t}}(\cdot)$ 和 $\mathbf{g}_{\sigma_{z_t}}(\cdot)$ 参数化, 网络的输入为 $t-1$ 时刻随机变量 \mathbf{z}_{t-1} 采样值; \mathbf{z}_0 为随机初始化的向量。神经网络 $\mathbf{g}_{\mu_{z_t}}(\cdot)$ 和 $\mathbf{g}_{\sigma_{z_t}}(\cdot)$ 计算过程如式(16)~式(20)所示。

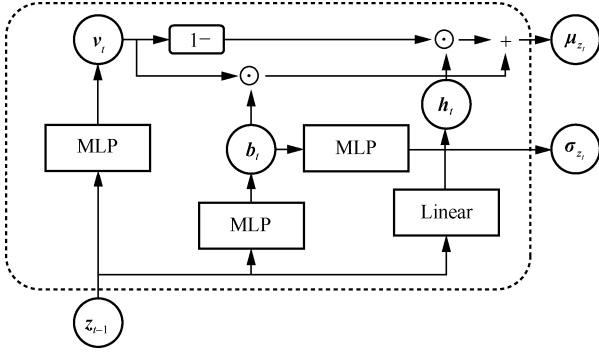


图 4 GTF 网络结构

$$\mathbf{v}_t = \text{MLP}(z_{t-1}, \text{ReLU}, \text{Sigmoid}) \quad (16)$$

$$\mathbf{b}_t = \text{MLP}(z_{t-1}, \text{ReLU}) \quad (17)$$

$$\mathbf{h}_t = \text{Linear}(z_{t-1}) \quad (18)$$

$$\mathbf{g}_{\mu_{z,t}}(z_{t-1}) = (1 - \mathbf{v}_t) \odot \mathbf{h}_t + \mathbf{v}_t \odot \mathbf{b}_t \quad (19)$$

$$\mathbf{g}_{\sigma_{z,t}}(z_{t-1}) = \text{MLP}(\text{ReLU}(\mathbf{b}_t), \text{Softplus}) \quad (20)$$

式(16)表示具有 ReLU 和 Sigmoid 激活函数的 MLP，式(17)表示具有 ReLU 激活函数的 MLP，式(18)表示 Linear 层，式(20)表示具有 ReLU 和 Softplus 激活函数的 MLP。

2.4 优化目标

ST-MTS-AD 模型的优化目标为最大化式(21)所示的证据下界 (ELBO, evidence lower bound)。

$$\mathcal{L}(\mathbf{x}_{1:T}; \phi, \theta) = \mathbb{E}_{q_\phi(z_{1:T} | \mathbf{x}_{1:T})} \left[\sum_{t=1}^T [\ln p_\theta(\mathbf{x}'_t | z_t) - D_{\text{KL}}(q_\phi(z_t | z_{t-1}, \mathbf{x}_{1:T}) \| p_\theta(z_t | z_{t-1}))] \right] \quad (21)$$

ST-MTS-AD 模型将重构概率对数似然 $\log p_\theta(\mathbf{x}'_t | z_t)$ 作为测试样本 $\mathbf{x}_{1:T}$ 中每个观测变量的异常得分。具体异常检测阈值选取方法如下。依次将测试集中每个样本的观测变量 \mathbf{x}_t 的重构概率对数似然值作为异常检测阈值，计算相应的 F1 得分，最大 F1 分数对应的重构概率似然值被选为测试样本 $\mathbf{x}_{1:T}$ 的异常检测阈值。如果 $\mathbf{x}_{1:T}$ 中观测变量 \mathbf{x}_t 的重构概率似然值低于阈值，则被标记为异常；否则被标记为正常。具体训练算法如算法 1 所示。

算法 1 ST-MTS-AD 模型的训练算法

输入 预处理的训练集 $D(\mathbf{x}_{1:T})$

输出 训练网络 f 、 g 、 ψ

- 1) 初始化随机变量 z_0
- 2) 循环
- 3) 从训练集中采样一个 batch 的样本 $\mathbf{x}_{1:T} \sim P_D(\mathbf{x}_{1:T})$
- 4) $\mathcal{L}_{\text{kl}} \leftarrow 0, \mathcal{L}_{\text{lh}} \leftarrow 0$

5) for $t=1, \dots, T$

6) 通过式(11)和式(15)计算 KL 散度 $\mathcal{L}_{\text{kl}} \leftarrow \mathcal{L}_{\text{kl}} + D_{\text{KL}}(\mathcal{N}(\psi_{\mu_{z,t}}(z_{t-1}, \mathbf{e}_t), \text{diag}(\psi_{\sigma_{z,t}}(z_{t-1}, \mathbf{e}_t))) \| \mathcal{N}(\mathbf{g}_{\mu_{z,t}}(z_{t-1}), \text{diag}(\mathbf{g}_{\sigma_{z,t}}(z_{t-1}))))$

7) $\xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ，按重采样过程得到 z_t 的采样值 $z_t \leftarrow \xi \odot \text{diag}(\psi_{\sigma_{z,t}}(z_{t-1}, \mathbf{e}_t)) + \psi_{\mu_{z,t}}(z_{t-1}, \mathbf{e}_t)$

8) 通过式(13)计算对数似然 $\mathcal{L}_{\text{lh}} \leftarrow \mathcal{L}_{\text{lh}} + \ln[\mathcal{N}(\mathbf{x}'_t | f_{\mu_x}(z_t, \mathbf{e}_t), \text{diag}(f_{\sigma_x}(z_t, \mathbf{e}_t)))]$

9) end for

10) $\mathcal{L}(\mathbf{x}_{1:T}; \phi, \theta) \leftarrow \mathcal{L}_{\text{lh}} - \mathcal{L}_{\text{kl}}$

11) 由式(21)采用 Adam 优化器更新参数

$$\theta, \phi \leftarrow \text{Adam}(-\nabla_{\theta, \phi} \mathcal{L}(\mathbf{x}_{1:T}; \phi, \theta))$$

12) 直到算法收敛

3 实验

3.1 实验数据集及环境

实验采用以下 4 个公开的数据集：SMD (server machine dataset) 来自一家大型互联网公司为期 5 周的服务器检测数据集^[2]；MSL (mars science laboratory) 和 SMAP (soil moisture active passive satellite) 来自 NASA 航天器监测系统报告中的遥感数据^[12]；SWaT (secure water treatment) 来自一个水处理厂信息物理系统为期 11 天的监控数据集^[3]。各数据集具体的描述如表 1 所示。

表 1	各数据集具体的描述		
数据集	训练数据长度	测试数据长度	维度
SMD	708 405	708 420	38
SMAP	135 183	427 617	25
MSL	58 317	73 729	55
SWaT	99 360	89 984	51

实验硬件环境如下：Ubuntu16.04 操作系统，Inter Xeon Gold 5220R CPU，NVIDIA Tesla T4 显卡。软件环境如下：Python3.6，Pytorch 1.10.1。

3.2 实验模型参数设置和评价指标

ST-MTS-AD 模型在 SMD、SMAP、MSL、SWaT 数据集上的滑动窗口大小 w 、滑动步幅 l 、观测序列长度 T 和自注意力头数 H 分别设为 10、10、200 和 8。根据数据集的维度大小，模型在 SMD 和 SWaT

数据集上的 Transformer 编码器输出维度 d 分别设为 128 和 256, 在 SMAP 和 MSL 数据集中设为 64。实验使用 Adam 优化器, 设置学习率 $r=0.000 1$, 批处理大小 $batch=64$, 训练周期数 $epoch=200$ 。

ST-MTS-AD 目标是检测观测序列 $\mathbf{x}_{1:T}$ 中观测变量 \mathbf{x}_i 是否异常。在测试集中, 若观测变量 \mathbf{x}_i 表示的滑动窗口中某一时间点取值 $k_i \in \mathbb{R}^M$ 为异常点, 则 \mathbf{x}_i 标记为异常。对于测试集中的每个观测变量 \mathbf{x}_i , 如果重构出的观测变量 \mathbf{x}'_i 被判断为异常, 且 \mathbf{x}_i 真实标签也为异常, 则记为真阳 (TP)。如果重构出的观测变量 \mathbf{x}'_i 被判断为异常, 但 \mathbf{x}_i 真实标签为正常, 则记为假阳 (FP)。如果重构出的观测变量 \mathbf{x}'_i 被判断为正常, 但 \mathbf{x}_i 真实标签为异常, 则记为假阴 (FN)。如果重构出的观测变量 \mathbf{x}'_i 被判断为正常, 且 \mathbf{x}_i 真实标签为正常, 则记为真阴 (TN)。本文使用 3 个指标来衡量异常检测模型的性能, 分别为精确率 Precision、召回率 Recall、F1 分数, 其中, F1 分数为精确率和召回率的调和平均数, F1 分数越大表示异常检测模型的性能越好。

$$\begin{cases} \text{Precision} = \frac{TP}{TP+FP} \\ \text{Recall} = \frac{TP}{TP+FN} \\ \text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{cases} \quad (22)$$

3.3 实验结果与分析

选取 6 种与 ST-MTS-AD 相关的异常检测模型进行实验对比, 分别为 DAGMM^[6]、LSTM-VAE^[8]、MSCRED^[15]、USAD (unsupervised anomaly detec-

tion)^[16]、OmniAnomaly^[2]、SDFVAE^[9]。各模型的实验均采用本文的数据预处理方式, 不同模型的性能对比如表 2 所示。

从表 2 可知, ST-MTS-AD 模型在 SMD、SMAP、MSL 和 SWaT 数据集上的 F1 分数分别为 0.933 2、0.966 4、0.981 9 和 0.834 2, 相较于 5 种对比模型整体上有较高的提升。在 SMD、SMAP、MSL 和 SWaT 数据集上, ST-MTS-AD 模型的 F1 分数比 MSCRED 分别提高了 15.6%、10.0%、4.7%和 3.3%, ST-MTS-AD 模型的 F1 分数比 USAD 模型分别提高了 9.6%、6.9%、15.6%和 3.0%, 与 MSCRED 相比, ST-MTS-AD 在 SMD 数据集上的 F1 分数提高最多, 这是因为 SMD 数据集中存在持续时间短、异常偏差较小的取值片段^[9], MSCRED 模型中的构造签名矩阵无法捕获这些细微的异常特征。USAD 虽然采用对抗学习策略避免自编码器无法通过重构误差区分正常样本与异常样本的问题, 但其自编码器网络没有捕获 MTS 的时序依赖特征。实验结果也验证了基于重构的随机性模型 ST-MTS-AD 的异常检测效果好于基于重构的确定性模型 MSCRED 和 USAD。

DAGMM 的 F1 分数在 SMD、SMAP、MSL 和 SWaT 数据集上比 ST-MTS-AD 模型分别低 8.4%、7.9%、17.7%和 2.3%, DAGMM 虽属于基于重构的随机性模型, 但其自编码器网络结构设计面向表格数据, 无法提取 MTS 时序特征。ST-MTS-AD 模型在 SMD、SMAP 和 MSL 数据集上的 F1 分数比 SDFVAE 分别提高了 3.3%、7.8%和 12.4%, 但在 SWaT 数据集上的 F1 分数比 SDFVAE 低 1.8%, 这是由于 SDFVAE 比 ST-MTS-AD 模型有较强的抗噪能力。LSTM-VAE

表 2 ST-MTS-AD 模型与 6 种模型的性能对比

模型	SMD			SMAP			MSL			SWaT		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
DAGMM	0.879 0	0.832 5	0.855 1	0.800 8	0.999 9	0.889 4	0.676 8	0.998 9	0.807 2	0.997 7	0.687 8	0.814 3
LSTM-VAE	0.974 1	0.804 6	0.881 3	0.997 1	0.888 5	0.939 7	0.995 1	0.847 6	0.915 5	0.998 3	0.684 2	0.812 0
MSCRED	0.682 9	0.986 7	0.807 2	0.784 0	0.990 1	0.878 9	0.883 1	0.999 9	0.937 9	0.999 2	0.677 0	0.807 2
USAD	0.870 9	0.832 5	0.851 3	0.825 6	0.999 9	0.904 4	0.796 9	0.909 9	0.849 1	0.969 6	0.695 6	0.810 1
OmniAnomaly	0.873 8	0.844 0	0.858 6	0.764 8	0.999 9	0.866 7	0.890 2	0.863 7	0.876 7	0.988 6	0.677 0	0.803 6
SDFVAE	0.882 1	0.926 3	0.903 0	0.884 6	0.908 0	0.896 0	0.853 4	0.894 5	0.873 0	0.944 8	0.772 1	0.849 7
ST-MTS-AD	0.893 1	0.977 2	0.933 2	0.998 6	0.936 2	0.966 4	0.993 7	0.970 4	0.981 9	0.975 8	0.728 4	0.834 2

和 OmniAnomaly 为与 ST-MTS-AD 最相关的模型，均为基于 VAE 的序列生成模型。在 SMD、SMAP、MSL 和 SWaT 数据集上，ST-MTS-AD 模型的 F1 分数比 LSTM-VAE 模型提高了 5.9%、2.8%、7.3% 和 2.7%，比 OmniAnomaly 模型提高了 8.6%、11.5%、11.9% 和 3.8%。这是由于 LSTM-VAE 模型无法建模隐空间中随机变量之间的时序依赖性。OmniAnomaly 模型基于 RNN 各时刻的隐向量实现随机变量间时序依赖性，这种机制无法实现随机变量间的长时依赖关联，其采用的基于 LGSSM 的随机变量先验分布生成方式无法实现随机变量间的非线性转换，且该模型在生成网络中仅依赖各时刻随机变量的采样值，没有利用推断网络 RNN 的隐向量信息。

3.4 消融实验

为了验证 ST-MTS-AD 模型相关模块设计的有

效性，将 ST-MTS-AD 模型与其 3 个变体进行对比，3 个变体分别为 ST-MTS-AD-1、ST-MTS-AD-2 和 ST-MTS-AD-3。ST-MTS-AD-1 模型表示在图 1 的 ST-MTS-AD 模型基础上将推断网络中 Transformer 编码器替换为门控循环单元（GRU, gate recurrent unit）神经网络。ST-MTS-AD-2 模型表示将 ST-MTS-AD 模型中用于生成随机变量先验分布的 GTF 替换为 LGSSM。ST-MTS-AD-3 模型表示图 1 生成网络去除了 e_t 作为输入的设计，生成网络的输入仅来自随机变量 z_t 的采样值。各模型在 4 个数据集上的实验结果如图 5 所示。

从图 5 可知，ST-MTS-AD 模型在 SMD、SMAP、MSL 和 SWaT 数据集上的 F1 分数比 ST-MTS-AD-1 模型分别提高了 11.1%、2.6%、7.3% 和 6.3%，因此基于 Transformer 编码器生成的时序依赖特征能更好地实现隐空间中随机变量间的时序依赖性。

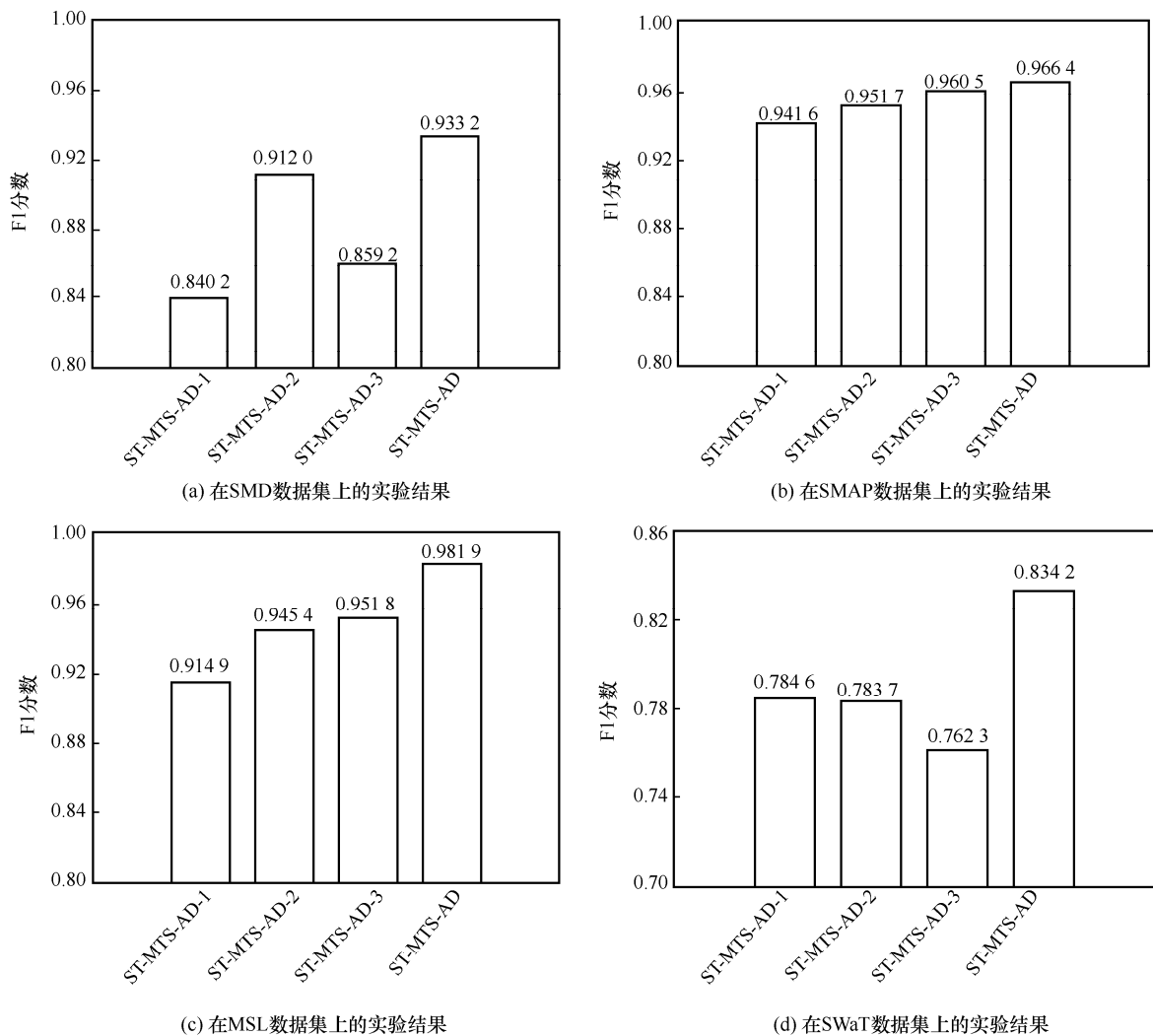


图 5 不同数据集上的消融实验

ST-MTS-AD 模型的 F1 分数比 ST-MTS-AD-2 模型分别提高了 2.3%、1.5%、3.9%和 6.4%，原因是 GTF 采用 GRU 的思想对随机变量 z_{t-1} 到 z_t 的转换函数进行参数化，通过非线性转换函数控制 z_{t-1} 到 z_t 的信息传递，可以捕获随机变量之间更复杂的依赖性，而 LGSSM 利用卡尔曼滤波的思想实现随机变量间的线性转换，结果证明了 ST-MTS-AD 使用 GTF 的连接方式比使用 LGSSM 的连接方式更加有效。ST-MTS-AD 模型的 F1 分数在 SMD、SMAP、MSL 和 SWaT 数据集分别比 ST-MTS-AD-3 模型提高了 8.6%、0.6%、3.2%和 9.4%，由 Transformer 编码器生成时依赖特征 e_t 和推断网络生成的随机变量 z_t 的采样值能更好地重构 MTS 各时刻 x'_t 的分布。另外，与 ST-MTS-AD-1、ST-MTS-AD-2 和 ST-MTS-AD-3 相比，ST-MTS-AD 在 SMAP 数据集上的 F1 分数提高并不明显。这是因为 SMAP 数据集中存在很多离散变量，其异常片段比较容易被检测。

4 结束语

本文提出了一种融合 Transformer 编码器和 VAE 的随机 Transformer MTS 异常检测模型。该模型基于 Transformer 编码器生成的时序特征实现隐空间中随机变量间的长时依赖性，采用门控转换函数生成时序随机变量的先验分布，由推断网络生成的各时刻随机变量近似后验分布采样值和 Transformer 编码器输出的时序特征重构 MTS 各时刻取值的分布。在 4 个公开数据集上实验结果表明了 ST-MTS-AD 设计的有效性。下一步笔者将研究如何基于 Transformer 编码器实现隐空间中随机变量间的非马尔可夫动态性。

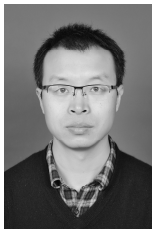
参考文献:

- [1] 丁小欧, 于晟健, 王沐贤, 等. 基于相关性分析的工业时序数据异常检测[J]. 软件学报, 2020, 31(3): 726-747.
DING X O, YU S J, WANG M X, et al. Anomaly detection on industrial time series based on correlation analysis[J]. Journal of Software, 2020, 31(3): 726-747.
- [2] SU Y, ZHAO Y J, NIU C H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2019: 2828-2837.
- [3] LI D, CHEN D, JIN B, et al. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks[C]//International Conference on Artificial Neural Networks. Berlin: Springer, 2019: 214-232.
- [4] 黄训华, 张凤斌, 樊好义, 等. 基于多模态对抗学习的无监督时间序列异常检测[J]. 计算机研究与发展, 2021, 58(8): 1655-1667.
HUANG X H, ZHANG F B, FAN H Y, et al. Multimodal adversarial learning based unsupervised time series anomaly detection[J]. Journal of Computer Research and Development, 2021, 58(8): 1655-1667.
- [5] 段雪源, 付钰, 王坤. 基于 VAE-WGAN 的多维时间序列异常检测方法[J]. 通信学报, 2022, 43(3): 1-13.
DUAN X Y, FU Y, WANG K. Multi-dimensional time series anomaly detection method based on VAE-WGAN[J]. Journal on Communications, 2022, 43(3): 1-13.
- [6] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection[C]//International Conference on Learning Representations. Piscataway: IEEE Press, 2018: 1-19.
- [7] DENG L W, CHEN X H, ZHAO Y, et al. HIFI: anomaly detection for multivariate time series with high-order feature interactions[C]//International Conference on Database Systems for Advanced Applications. Berlin: Springer, 2021: 641-649.
- [8] 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.
- [9] DAI L, LIN T, LIU C, et al. SDFVAE: static and dynamic factorized VAE for anomaly detection of multivariate CDN KPIs[C]//Proceedings of the Web Conference 2021. New York: ACM Press, 2021: 3076-3086.
- [10] LI L Y, YAN J C, WANG H Y, et al. Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(3): 1177-1191.
- [11] AHMAD S, LAVIN A, PURDY S, et al. Unsupervised real-time anomaly detection for streaming data[J]. Neurocomputing, 2017, 262: 134-147.
- [12] HUNDMAN K, CONSTANTINOU V, LAPORTE C, et al. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 387-395.
- [13] ZHANG Y X, CHEN Y Q, WANG J D, et al. Unsupervised deep anomaly detection for multi-sensor time-series signals[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(2): 2118-2132.
- [14] MALHOTRA P, RAMAKRISHNAN A, ANAND G, et al. LSTM-based encoder-decoder for multi-sensor anomaly detection[J]. arXiv Preprint, arXiv:1607.00148, 2016.
- [15] ZHANG C X, SONG D J, CHEN Y C, et al. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time

series data[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 1409-1416.

- [16] AUDIBERT J, MICHIARDI P, GUYARD F, et al. USAD: unsupervised anomaly detection on multivariate time series[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2020: 3395-3404.
- [17] KRISHNAN R G, SHALIT U, SONTAG D. Structured inference networks for nonlinear state space models[C]//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. New York: ACM Press, 2017: 2101-2109.
- [18] CHUNG J, KASTNER K, DINH L, et al. A recurrent latent variable model for sequential data[J]. Advances in Neural Information Processing Systems, 2015, 28: 2962-2970.
- [19] FRACCARO M, SNDRBY S K, PAQUET U, et al. Sequential neural models with stochastic layers[J]. Advances in Neural Information Processing Systems, 2016, 29: 2199-2207.
- [20] ZERVEAS G, JAYARAMAN S, PATEL D, et al. A transformer-based framework for multivariate time series representation learning[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2021: 2114-2124.

[作者简介]



霍纬纲（1978—），男，山西洪洞人，博士，中国民航大学教授、硕士生导师，主要研究方向为时序大数据分析挖掘。



梁锐（1999—），男，安徽淮南人，中国民航大学硕士生，主要研究方向为数据挖掘、多维时间序列异常检测。



李永华（1974—），女，内蒙古赤峰人，中国民航大学讲师，主要研究方向为数据挖掘。