时空变分图注意力自编码器用于工业故障检测

胥舒元
成都理工大学　四川成都　610059

摘　要：在工业物联网、数据中心运维和金融风控等关键领域，对多变量时间序列（MTS）进行精确的异常检测是保障系统可靠性和安全性的核心需求。MTS 中的异常往往并非孤立的数值偏差，而是体现为多个相互关联的变量之间正常协同模式的微妙破坏。具体而言，其编码器虽能感知图结构，解码器却往往简化为结构盲的多层感知机（MLP），导致在重构过程中丢失了宝贵的结构信息，从而限制了对“关系异常”的检测精度。

关键词：工业故障检测；多变量时间序列检测；编码器

1. 引言

1.1 研究背景与问题重要性

我们正处在一个由数据驱动的时代，从为全球经济提供动力的工业生产线，到支撑现代信息社会运转的大型数据中心，再到保障人类探索深空的航天器，这些复杂系统的稳定运行都依赖于海量传感器网络的实时监控。这些传感器网络持续不断地产生高维度的多变量时间序列（Multivariate Time Series, MTS）数据，它们如同系统的“生命体征”，精确记录着从温度、压力、网络流量到 CPU 负载等各项关键指标的动态变化。因此，开发能够自动、实时、准确地从这些 MTS 数据中检测出异常（Anomaly）的技术，已不再是一个单纯的学术问题，而是保障现代社会基础设施安全、预防灾难性故障、避免重大经济损失的核心技术需求。

1.2 现有方法及其局限性

为了应对这一挑战，学术界和工业界均付出了巨大努力。传统的统计方法，如主成分分析（PCA）或向量自回归（VAR），因其对数据分布的严格假设和线性关系的局限性，难以适应现代 MTS 数据的高度非线性和动态性。近年来，以深度学习为代表的数据驱动方法，凭借其强大的非线性建模能力，已成为该领域的研究主流。这些方法大多遵循无监督的学习范式，通过在大量正常数据上训练，学习一个能够精确描述“正常模式”的模型，进而将无法被模型很好解释的数据点识别为异常 [1]。

现有深度学习方法大致可以沿着两条技术路线演进。第一条路线侧重于捕捉 MTS 中复杂的时间依赖性。以 Malhotra 等人 [cite] 开创性地使用 LSTM 自编码器（LSTM-AE）为代表，这类方法通过循环神经网络（RNN）及其变体（如 LSTM、GRU）来学习和重构时间序列。后续的工作如经典的 OmniAnomaly [cite Su, et al.]，通过将 RNN 与变分自编码器（VAE）和更复杂的随机过程（如归一化流）相结合，极大地增强了模型对时间序列概率分布的建模能力。

第二条路线则敏锐地抓住了变量间关系的重要性，开创性地引入了图神经网络（GNN）。以 GDN [cite Zhao, et al.] 和 STG-AE [cite Deng & Hooi] 为代表的工作，将 MTS 中的 N 个变量建模为图的 N 个节点，通过预先计算或从数据中学习变量间的关系作为边，然后利用 GNN 来同时聚合时间和空间（结构）维度的信息。

2. 背景知识

2.1 多变量时间序列异常检测

多变量时间序列 (MTS) X = {x_1, x_2, ..., x_T} 是一个由 T 个时间戳组成的数据序列，其中每个时间戳 x_t ∈ R^N 是一个 N 维的向量，代表在 t 时刻 N 个不同变量（或传感器）的观测值。MTS 异常检测的目标是识别出序列中与正常行为模式不符的时间点或时间段。根据异常的类型，可以分为点异常、上下文异常和群体（关系）异常。我们的工作主要关注后两者，特别是由于变量间协同关系破坏导致的群体异常。

2.2 变分自编码器

变分自编码器（VAE）是由 Kingma 和 Welling [cite] 提出的深度生成模型，它在标准自编码器（AE）的基础上引入了概率论的视角，旨在学习数据的潜在概率分布。一个 VAE 由两部分组成：

编码器 (Encoder): 也称为推断网络或识别网络 q_

ϕ(z|x)。它接收输入数据 x，并输出一个潜在分布的参数，通常是高斯分布的均值 μ 和对数方差 log(σ^2)。它学习的是从数据空间到潜在空间的映射。

解码器 (Decoder): 也称为生成网络 p_θ(x|z)。它从通过 μ 和 log(σ^2) 参数化的潜在分布中采样一个潜在向量 z（通过重参数化技巧），然后尝试从 z 重构出原始的输入数据 x'。它学习的是从潜在空间到数据空间的映射 [2]。

VAE 的优化目标是最大化证据下界 (Evidence Lower Bound, ELBO)，这等价于最小化一个包含两项的损失函数：

Loss_VAE = Loss_reconstruction + β * Loss_KL

其中，Loss_reconstruction（如均方误差 MSE）衡量了重构数据 x' 与原始数据 x 的相似度。Loss_KL（KL 散度）衡量了编码器产生的潜在分布 q_ϕ(z|x) 与一个预设的先验分布 p(z)（通常是标准正态分布 N(0, I)）之间的差异。β 是平衡两者的超参数。KL 散度项起到了正则化的作用，它迫使编码器学习一个结构良好、连续且完整的潜在空间，这不仅增强了模型的泛化能力，也使其能够生成新的、与训练数据相似的样本。

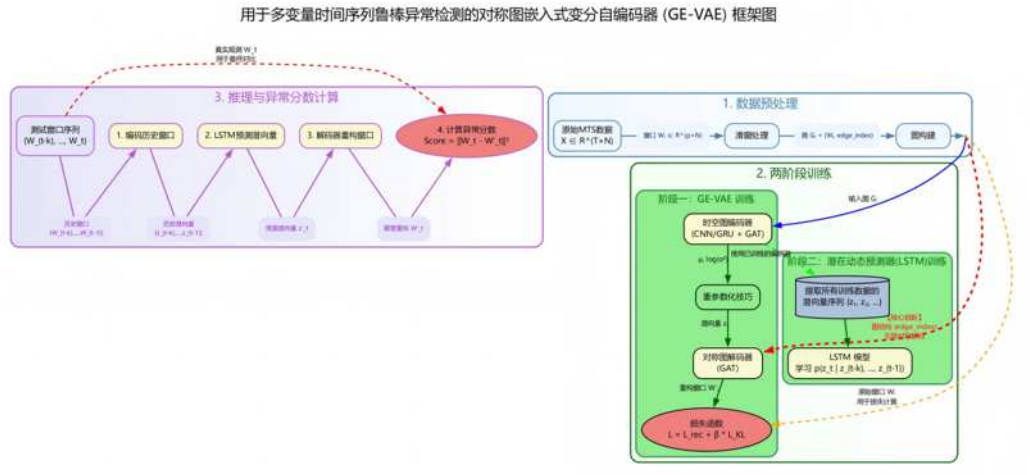
3. 模型与方法

3.1 动机

我们工作的核心动机源于对现有基于自编码器的 MTS 异常检测方法中一个普遍存在的“非对称性”瓶颈的观察。许多先进的方法已经认识到利用 GNN 来编码变量间的空间依赖关系的重要性，然而，它们的解码过程往往被简化为一个从全局潜在向量到高维原始数据的直接映射，通常由一个结构盲的 MLP 完成。我们认为，这种设计存在一个根本性的缺陷：它在编码阶段辛苦学习到的、细粒度的、节点级的结构化信息，在图池化（Graph Pooling）操作后被压缩进一个全局向量，解码器却无法再利用原始的图拓扑来引导这个“解压缩”过程。

3.2 模型总体架构

我们的 GE-VAE 框架遵循一个两阶段的、基于“预测性重构”的异常检测范式。其总体流程如下图所示：



用于多变量时间序列鲁棒异常检测的对称图嵌入式变分自编码器 (GE-VAE) 框架图

数据预处理：原始的 MTS 数据 X ∈ R^(T × N) 首先经过滑窗处理，被切分为一系列重叠的窗口 W_i ∈ R^(p × N)，其中 p 是窗口长度。每个窗口 W_i 被视为一个图 G_i，其中 N 个变量是节点，节点间的关系通过预定义的图构建方法（如相关性）确定，形成边索引 edge_index。

阶段一：GE-VAE 训练：

编码：每个图窗口 G_i 被送入时空图编码器，该编码器通过 CNN/GRU 和 GAT 层，将其压缩成一个潜在分布的参数 μ_i 和 log(σ^2)_i。

解码：从该分布中采样一个潜在向量 z_i，然后送入对称图解码器。解码器利用 edge_index，通过 GAT 层进行结构化的信息扩展，最终重构出窗口 W'_i。

优化：模型通过最小化节点重构损失和 KL 散度损失进行训练，学习如何精确地压缩和解压“正常”的图窗口。

阶段二：LSTM 训练：

所有训练集的图窗口通过训练好的 GE-VAE 编码器，生成一个潜在向量的时间序列 {z_1, z_2, ..., z_M}。

潜在动态预测器 (LSTM) 在这个 z 序列上进行训练，学

习预测下一时刻的潜在向量，即学习 p(z_t | z_{t−k}, ..., z_{t−1})。

推理与异常分数计算 :

对于测试集中的一个新序列 {W_{t−k}, ..., W_t}，我们首先将其前 k 个窗口 {W_{t−k}, ..., W_{t−1}} 送入 GE−VAE 编码器，得到潜在向量 {z_{t−k}, ..., z_{t−1}}。

将这 k 个历史 z 向量送入训练好的 LSTM，预测出下一时刻的期望潜在向量 z’_t。

将这个预测出的 z’_t 送入 GE−VAE 的图解码器，重构出期望的窗口 W’_t。

计算期望重构窗口 W’_t 与真实观测窗口 W_t 之间的差异（如平方误差），该误差即为 t 时刻的异常分数。如果真实观测与基于历史的期望严重不符，则异常分数会很高 [3]。

3.3 时空图编码器

编码器的目标是将一个高维的、结构化的图窗口 G_i = (X_i, A)，其中 X_i ∈ R^(N × p) 是节点特征矩阵（每个节点的 p 维时间序列），A 是邻接矩阵（由 edge_index 表示），映射到一个低维的潜在分布 q_ϕ(z_i|G_i)。这个过程分为两步：

节点级时间特征提取 : 对于每个节点（变量），其 p 维的时间序列首先需要被编码成一个固定维度的特征向量。我们探索了两种方法：

CNN: 使用一维卷积网络（1D−CNN）来捕捉时间序列中的局部模式和形状特征。

GRU: 使用门控循环单元（GRU）来捕捉序列的顺序依赖关系。

经过这一步，每个节点的 p 维时间序列被转换为一个 d_t 维的初始节点嵌入 h_j^0 ∈ R^(d_t)。

图级空间信息聚合 : 随后，所有节点的初始嵌入 H^0 = {h_1^0, ..., h_N^0} 和图的边索引 edge_index 被送入一个两层的图注意力网络（GAT）：

H^(l+1) = GAT(H^l, edge_index)

GAT 通过消息传递和自注意力机制，让每个节点的表示能够融入其邻居的信息。经过两层 GAT 后，我们得到最终的节点表示 H^L，它同时编码了每个变量自身的时间动态和变量间的空间依赖关系。

图池化与潜在分布输出 : 为了得到代表整个图窗口的单一向量，我们对所有最终节点表示 H^L 进行一次全局平均池化（Global Mean Pooling）操作：

h_G = GlobalMeanPool(H^L)

最后，这个图级别的表示 h_G 通过两个独立的线性层，分别输出潜在分布的均值 μ 和对数方差 log(σ^2)：

μ = FC_μ(h_G)

log(σ^2) = FC_logvar(h_G)

3.4 对称图解码器

这是我们方法论的核心。解码器的任务是执行编码器的“逆操作”：给定一个从潜在分布中采样得到的全局向量 z ∈ R^(d_z)，以及图的结构信息 edge_index，重构出原始的节点特征 X ∈ R^(N × p)。

广播 / 逆池化 (Broadcast/Un−pooling): 这是图解码的第一步，也是与 MLP 解码器最本质的区别。我们将代表全局信息的潜在向量 z 复制 N 份，为图中的每一个节点都分配一个初始表示。更形式化地说，对于一个批次中的所有图，我们利用 batch_map 索引将 z 广播到对应的节点上，得到初始节点表示 H_dec^0。

H_dec^0 = FC_initial(z[batch_map])

其中 FC_initial 是一个线性层，用于将 z 的维度匹配到 GNN 的隐藏维度。

结构化信息扩展 : 随后，我们将 H_dec^0 和 edge_index 送入一个与编码器结构对称的两层 GAT 网络 [4]。

H_dec^(l+1) = GAT_dec(H_dec^l, edge_index)

在这个过程中，GNN 扮演了“信息扩散器”和“关系协调器”的角色。每个节点的表示不再是孤立的，而是通过消息传递不断地与邻居进行“协商”，根据图结构来调整和细化自身的表示，以确保最终重构出的所有节点特征能够符合它们之间应有的协同关系。

4. 实验

4.1 数据集与评估指标

数据集 : 我们在公开的、由 Su 等人 [cite] 发布的服务器监控数据集 SMD (Server Machine Dataset) 上验证我们的方法。我们选取了其中的 machine−1−1 子集进行实验，该子集包含 38 个传感器变量，记录了 28,479 个训练时间点和 28,479 个测试时间点。这是一个典型的高维、复杂 MTS 数据集。

评估指标 : 我们采用时间序列异常检测领域标准的评估指标：精确率（Precision）、召回率（Recall）和 F1 分数

（F1–Score）。我们遵循 point–adjust 的评估策略 [cite Su, et al.]，即只要真实异常段内的任何一个点被检测到，则整个段内的所有点都被视为正确检测 [5]。

4.2 实现细节与超参数

我们的模型完全基于 PyTorch 和 PyTorch Geometric 实现。关键超参数设置如下：滑窗长度 l_win=48，LSTM 序列长度 l_seq=12，潜在空间维度 code_size=32，批大小 batch_size=32。GE–VAE 和 LSTM 均训练 100 个周期，学习率分别为 0.0006 和 0.0003。图构建采用皮尔逊相关系数法，阈值为 0.8。所有实验均在单个 NVIDIA GeForce RTX 3080 GPU 上完成。

随机种子 (Seed)	Precision	Recall	F1 Score
42	0.6601	0.9165	0.7674
123	0.6325	0.8781	0.7353
1024	0.6268	0.8703	0.7288
平均值 (Mean)	0.6398	0.8883	0.7438
标准差 (Std)	0.0146	0.0205	0.0171

最终，我们的 GE–VAE 模型可以被科学地报告为取得了 0.744 ± 0.017 的平均 F1 分数。较低的标准差证明了我们模型的稳健性。即使在最差的一次随机运行中（F1=0.7288），其性能也处于非常高的水平。而其最佳表现可达 F1=0.7674，展示了其巨大的潜力。

5. 结论

本文针对多变量时间序列异常检测中，现有方法对变量间协同关系建模不足以及编解码架构不对称的问题，提出了一种对称图嵌入式变分自编码器（GE–VAE）。结合 LSTM 对潜在空间动态的预测，我们的框架能够有效地通过预测性重构误差来识别异常。

在公开的 SMD 数据集上的大量实验证明了我们方法的有效性。特别地，我们通过严谨的多次独立实验和系统的后处理调优。我们的消融实验清晰地验证了对称图解码器的优越性，同时对引入边重构损失的失败探索也为领域内研究提供了宝贵的实践洞见。

参考文献：

[1] 黎国强，魏美容，吴德烽，等 . 零故障样本下小波知识驱动的工业机器人故障检测 [J]. 仪器仪表学报 ,2024,45(09):166–176

[2] 测试性验证试验中的故障样本综合加权分配方法 . 张西山 ; 黄考利 ; 敖贤野 ; 连光耀 ; 邱文昊 . 计算机测量与控制 ,2015(01)

[3] 高档数控机床故障样本获取方法与建库技术研究 . 王晗 ; 徐小力 ; 吴国新 . 机械工业标准化与质量 ,2011(04)

[4] 数字化和智能化赋能高职专业教学改革创新研究——以工业机器人技术专业为例 . 李萌 . 中国机械 ,2025(12)

[5] 职业本科视角下工业机器人技术专业人才培养模式探索 . 李晓慧 . 柳州职业技术大学学报 ,2025(03)

作者简介： 胥舒元，男，汉，2001.09.24，四川西充，硕士，智能仪器仪表