



# 基于堆叠式对抗变分循环神经网络的多维时间序列异常检测

陈文超<sup>1</sup>, 方博为<sup>1</sup>, 代良<sup>2</sup>, 陈渤<sup>1\*</sup>, 刘畅<sup>2</sup>, 赵小楠<sup>3\*</sup>

1. 西安电子科技大学雷达信号处理国家级重点实验室, 西安 710071

2. 中国科学院信息工程研究所, 中国科学院大学, 北京 100093

3. 山东新一代信息产业技术研究院有限公司, 济南 250100

\* 通信作者. E-mail: bchen@mail.xidian.edu.cn, zxnan891100214@163.com

收稿日期: 2022-07-11; 修回日期: 2022-09-16; 接受日期: 2022-11-18; 网络出版日期: 2023-09-11

国家自然科学基金 (批准号: 6220010437, U21B2006, 61771361)、雷达信号处理全国重点实验室基金 (批准号: JKW202X0X)、陕西省高校青年创新团队支持项目、111 引智基地项目 (批准号: B18039) 和中组部高层次海外人才引进计划资助

**摘要** 为了建模多维时间序列 (multivariate time series, MTS) 中复杂的时序依赖性与随机性, 并实现对它的无监督异常检测这一工业机器或互联网基础设施设备质量管理中的关键任务, 本文提出了一种堆叠式对抗变分循环神经网络 (stacked adversarial variational recurrent neural network, SaVRNN). SaVRNN 是一个层次化概率动态模型, 它将层次化概率生成模型、堆叠式循环结构和多层对抗优化方式整合到一个联合贝叶斯框架下. 具体来说, SaVRNN 核心思想是利用堆叠循环结构捕捉多层次与长距离的时序相关性, 利用层次化的概率生成操作建模多层的结构特性, 进而实现对多维时间序列正常模式的学习, 最后通过重构的概率来判断异常模式. 为了实现模型的高效推理, 本文创新性地提出了一种向上-向下对抗变分推理方案, 实现对隐层变量后验的准确估计. 针对多层对抗网络中难以实现生成器与判别器的更新平衡导致的 SaVRNN 训练困难的问题, 本文基于条件传输 (conditional transport, CT) 提出了一个新的优化方法. 最后, 基于 SaVRNN 的层次化结构, 本文将多层似然进行融合以改进传统的异常检测策略. 在两个公共数据集和一个实测数据集上显示所提方法在 F1-score 指标上实现了相比目前主流方法的更优性能, 证明了所提模型在时间序列在线异常检测任务上的有效性.

**关键词** 异常检测, 多维时间序列, 对抗变分贝叶斯, 变分循环神经网络, 条件传输

## 1 引言

多维时间序列的异常检测一直是信息科学领域的一个基础场景, 因为它对于工业设备或者互联网设施的设备管理十分关键, 例如大数据中心的服务器<sup>[1~3]</sup>、内容分发网络系统<sup>[4]</sup>等. 虽然基于统计

**引用格式:** 陈文超, 方博为, 代良, 等. 基于堆叠式对抗变分循环神经网络的多维时间序列异常检测. 中国科学: 信息科学, 2023, 53: 1750–1767, doi: 10.1360/SSI-2022-0277  
Chen W C, Fang B W, Dai L, et al. Stacked adversarial variational recurrent neural network for anomaly detection of multivariate time series (in Chinese). Sci Sin Inform, 2023, 53: 1750–1767, doi: 10.1360/SSI-2022-0277

的传统方法在一些时间序列异常检测场景中是有效的,例如单变量的时间序列<sup>[5~9]</sup>,它们在建模一些更加复杂与大量的多维时间序列场景时仍然缺乏有效性.因此,学术界与工业界将关注点转移到基于机器学习的异常检测方法,这些方法可以进一步分为有监督<sup>[8,10,11]</sup>和无监督<sup>[1,2,9,12~16]</sup>的异常检测.其中,有监督的异常检测模型经常遭受正常和异常样本不均衡的问题,因为在现实系统中缺少意外情况.因此,无监督的异常检测方法在近几年吸引到更多研究者的关心.无监督异常检测方法的核心挑战在于如何学习得到 MTS (multivariate time series) 数据的正常模式.在所有的无监督异常检测方法中,基于概率模型的方法同时考虑 MTS 数据中的时序依赖性与不确定性从而取得了优秀的检测性能.例如 OmniAnomaly<sup>[1]</sup>, SDFVAE<sup>[4]</sup> 等.尽管已有的概率动态模型在异常检测任务上已经比较高效,但是仍然存在两个问题可能限制它们的应用.一个是它们都是浅层模型,这限制了它们建模复杂动态特性的能力,并且无法抓捕长相关的时序依赖.另一个便是它们的隐层都被限制为具有对角协方差矩阵的高斯 (Gauss) 分布的形式,这降低了它们在学习正常模式下数据表示时的表现.

为了克服现有异常检测模型的不足,受到深度概率模型<sup>[17,18]</sup>和对抗训练方法<sup>[19]</sup>的启发,本文提出了一种新颖的深度概率动态模型,称之为堆叠式对抗变分循环神经网络 (stacked adversarial variational recurrent neural network, SaVRNN). SaVRNN 采用堆叠循环神经网络捕捉多层次与长相关的时序特性,采用具有层次化先验的概率生成模型增强模型对数据的适应能力,并使用对抗训练优化策略保证模型具有丰富的表征能力,进而将这些结构与策略融合成一个明确定义的变分生成模型.最后,将 SaVRNN 应用到 MTS 数据复杂结构和动态特性的建模中,并实现对它的在线异常检测<sup>[1,4]</sup>.

然而,作为一个具有层次化概率生成过程的多层对抗网络, SaVRNN 很难实现生成器和判别器更新的良好平衡,这使得它的训练十分困难,因此,迄今为止深层对抗生成模型很少被提出.鉴于这个问题,首先设计了一种向上-向下的自编码对抗变分推理方法,该方法将参数自下而上的似然信息与自上而下的先验信息进行整合,以实现准确的后验推理.然后,引进条件传输 (conditional transport, CT)<sup>[20~22]</sup>,它是一种量化两种概率分布差异的方法,用以修正损失函数以保证 SaVRNN 的收敛性.

本文的主要贡献总结如下.

(1) 本文提出一种能够通过捕捉多层次、长距离时序相关性而增强数据拟合能力的层次化概率动态模型 SaVRNN,并将其与新的异常检测策略结合应用于 MTS 在线异常检测任务.

(2) 为了对所提模型进行准确的对抗贝叶斯 (Bayes) 推理,本文提出一个有效的向上-向下自编码对抗推理方案.

(3) 基于 CT 理论提出一种改进的优化方法用以保证层次化对抗网络收敛.此外,为了充分利用所提深度模型得到的层次化表示,本文进一步改进了异常分数的计算方法.

(4) 通过大量的实验,本文证明所提 SaVRNN 在 F1-score 上明显优于目前的先进方法.

## 2 相关工作

学习 MTS 数据的正常模式是实现对其异常检测的关键步骤,近年来受到工程应用科学越来越多的关注.鉴于此,一些基于机器学习的方法被提出来.作为代表性研究, EncDec-AD<sup>[15]</sup> 使用基于 LSTM (long short-term memory) 的编码器解码器捕获 MTS 时间依赖性的正常模式,并根据重构误差确定异常. Telemanom<sup>[13]</sup> 利用 LSTM 预测航天器遥测通道的数值,并根据预测值与观测值之间的残差检测异常.此外, MSCRED<sup>[16]</sup> 引入一种多尺度卷积递归编码-解码器学习 MTS 中的空间相关性和时序特性,然后利用残差特征矩阵检测异常. MAD-GAN<sup>[14]</sup> 使用 LSTM-RNN 作为生成对抗网络 (generative adversarial network, GAN) 模型的基本框架来捕获正常的时空模式, OmniAnomaly<sup>[1]</sup> 引入随机递归神

神经网络 (stochastic RNN, SRNN) 帮助学习更鲁棒的表征. SDFVAE<sup>[4]</sup> 引入一种基于静态和动态因子化变分自编码的框架来显式地学习时不变和时变特性. OmniAnomaly 和 SDFVAE 都是利用重构似然检测异常, 因为似然越小, 越有可能为异常. 然而值得注意的是, 上述非监督工作均采用基于 RNN 的浅层结构来捕捉 MTS 中的时间依赖关系. 此外, 也有一些无监督异常检测方法忽略了 MTS 中不同时刻之间的时序依赖性. 例如, 无监督方法 DOMI<sup>[2]</sup> 将高斯混合变分自编码 (Gaussian mixture variational autoencoder, GmVAE) 与一维卷积相结合以检测大型数据中心的异常机器实例. DAGMM<sup>[23]</sup> 使用自编码器学习表征, 并使用高斯混合模型 (Gaussian mixture model, GMM) 进行分布的估计. 而 USAD<sup>[12]</sup> 设计了一种基于对抗训练的自编码器进行无监督的异常检测. 为了建模时间序列中的非平稳性, GmSVRNN<sup>[24]</sup> 将混合模型<sup>[2]</sup> 与开关机制<sup>[25~28]</sup> 引入到 VRNN<sup>[29]</sup>. 为了进一步考虑 MTS 中通道之间的相关性<sup>[30]</sup>, 一些基于图神经网络的无监督异常检测模型<sup>[31,32]</sup> 被提出来, 其中深度变分图卷积动态网络 (deep variational graph convolutional recurrent network, DVGCRN)<sup>[33]</sup> 创造性地将图网络结构融入动态概率生成网络中, 并实现了优秀的性能表现.

虽然现有的无监督异常检测方法在一些实际场景中非常有效, 但是一方面这些方法多为单层结构, 使得它们的表征能力受到限制; 另一方面虽然已经有深层概率动态建模方法实现了对 MTS 中时序和结构特征的建模, 但是它们的隐层往往限制为简单的分布形式, 无法适应复杂的数据分布场景. 面对该问题, 本文提出的 SaVRNN 是一种新颖的多层动态概率模型, 它不仅能够捕捉长距离的时序相关性, 而且能够考虑多层的结构特征, 同时继承了深度概率网络的各种优点, 并且隐式分布的隐变量与对抗训练方法保证了隐层的表征能力.

### 3 预备知识

本节首先对 VRNN 模型进行一个简短的概述, 它是本文工作的基础. 然后对无监督异常检测与分析任务的基本流程进行简单描述.

#### 3.1 变分循环神经网络

VRNN<sup>[29]</sup> 模型在每一个时间步都包含一个 VAE<sup>[34]</sup> 模型, 但与传统 VAE 不同, VRNN 的隐变量不再是一个标准的高斯分布, 而是服从以下分布:

$$\begin{aligned} \mathbf{z}_t &\sim \mathcal{N}(\boldsymbol{\mu}_{0,t}, \text{diag}(\boldsymbol{\sigma}_{0,t}^2)), \\ [\boldsymbol{\mu}_{0,t}, \boldsymbol{\sigma}_{0,t}] &= \varphi_{\tau}^{\text{prior}}(\mathbf{h}_{t-1}), \end{aligned} \quad (1)$$

其中  $\boldsymbol{\mu}_{0,t}$  和  $\boldsymbol{\sigma}_{0,t}$  分别表示条件先验分布的均值和方差参数. VRNN 的生成样本条件分布于随机隐变量  $\mathbf{z}_t$  和循环神经网络的隐状态  $\mathbf{h}_{t-1}$ , 具体如下:

$$\begin{aligned} \mathbf{x}_t | \mathbf{z}_t &\sim \mathcal{N}(\boldsymbol{\mu}_{x,t}, \text{diag}(\boldsymbol{\sigma}_{x,t}^2)), \\ [\boldsymbol{\mu}_{x,t}, \boldsymbol{\sigma}_{x,t}] &= \varphi_{\tau}^{\text{dec}}(\varphi_{\tau}^z(\mathbf{z}_t), \mathbf{h}_{t-1}), \end{aligned} \quad (2)$$

其中  $\boldsymbol{\mu}_{x,t}$  和  $\boldsymbol{\sigma}_{x,t}$  是生成分布的均值与方差参数, 它们的信息来自于  $\mathbf{z}_t$  和  $\mathbf{h}_{t-1}$ .  $\varphi_{\tau}^{\text{prior}}$ ,  $\varphi_{\tau}^{\text{dec}}$  和  $\varphi_{\tau}^z$  分别表示进行特征提取的不同的非线性映射函数, 例如神经网络. 为了推理得到  $\mathbf{z}_t$ , 和传统的 VAE 模型一样, 使用一个服从高斯分布的变分分布  $q(\mathbf{z}_t | \mathbf{x}_t)$  进行  $\mathbf{z}_t$  真实后验的估计. 与 VRNN 生成过程类似,

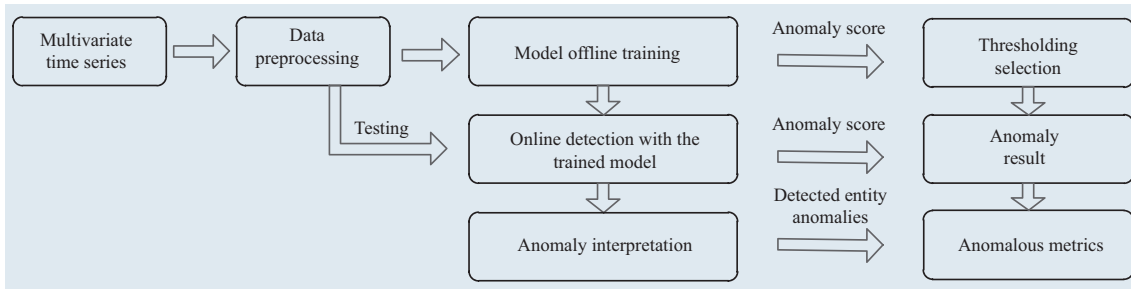


图 1 (网络版彩图) 无监督异常检测与分析的基本流程  
 Figure 1 (Color online) Procedure of unsupervised anomaly detection and diagnosis

推理过程变分分布被定义为  $\mathbf{x}_t$  和  $\mathbf{h}_{t-1}$  的函数:

$$\begin{aligned} \mathbf{z}_t | \mathbf{x}_t &\sim \mathcal{N}(\boldsymbol{\mu}_{z,t}, \text{diag}(\boldsymbol{\sigma}_{z,t}^2)), \\ [\boldsymbol{\mu}_{z,t}, \boldsymbol{\sigma}_{z,t}] &= \varphi_{\tau}^{\text{enc}}(\varphi_{\tau}^x(\mathbf{x}_t), \mathbf{h}_{t-1}), \end{aligned} \tag{3}$$

其中  $\boldsymbol{\mu}_{z,t}$  和  $\boldsymbol{\sigma}_{z,t}$  是变分分布的均值与方差参数. 在 VRNN 中, 变分分布的编码过程以及生成数据分布的解码过程通过动态隐状态  $\mathbf{h}_{t-1}$  结合在一起, 这使得数据中的时序相关性可以在不同时间步间进行传递. VRNN 的训练过程和传统 VAE 是一致的.

### 3.2 无监督异常检测与分析流程

图 1 展示了无监督异常检测的完整步骤. 可以看到, 对于采集得到的原始多维时间序列数据, 首先进行预处理以得到可以用来训练模型与测试的数据. 详细地, 使用了文献 [4] 中的归一化与时序滑动窗口方法. 然后, 对于所提无监督网络模型 SaVRNN 进行离线模型训练来学习正常模式的多维时间序列中的多层次复杂结构和动态特性. 之后, 基于 SaVRNN 的无监督重构似然作为异常检测的检测分数进行阈值的选取. 测试阶段, 对于新来的预处理之后的多维时间序列数据, 基于训练好的模型进行在线的异常检测, 得到检测结果. 最后可以类似文献 [4], 基于模型的异常分数进行异常的解释来确定异常的位置.

## 4 所提方法

本节首先给出本文所解决问题的定义并列出了完整框架. 然后, 针对 MTS 异常检测问题提出 SaVRNN 这个新颖的对抗层次化概率动态模型. 此外, 提出了一个向上 - 向下的变分推理方法实现所提模型的优化.

### 4.1 问题定义

首先将第  $n$  个 MTS 定义为  $\mathbf{x}_n = (\mathbf{x}_{1,n}, \mathbf{x}_{2,n}, \dots, \mathbf{x}_{T,n})$ , 这里  $n = 1, \dots, N$  且  $N$  是采集得到的 MTS 的数量.  $T$  是  $\mathbf{x}_n$  的持续时间,  $t$  时刻的观测向量为  $\mathbf{x}_{t,n} \in \mathbb{R}^V$ , 这里  $V$  表示某个时刻向量的维度, 因此  $\mathbf{x}_n \in \mathbb{R}^{T \times V}$ . MTS 数据的异常检测问题定义为: 某个时刻观测向量  $\mathbf{x}_{t,n}$  是否为异常值. 为了用无监督的方法有效地解决该问题, 需要一个强大的模型来学习输入数据的鲁棒表示.

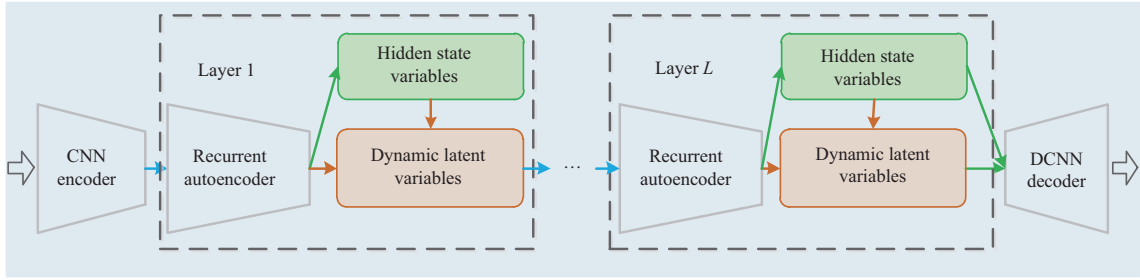


图 2 (网络版彩图) 本文提出的 SaVRNN 模型的整体架构图  
Figure 2 (Color online) Framework of the proposed SaVRNN

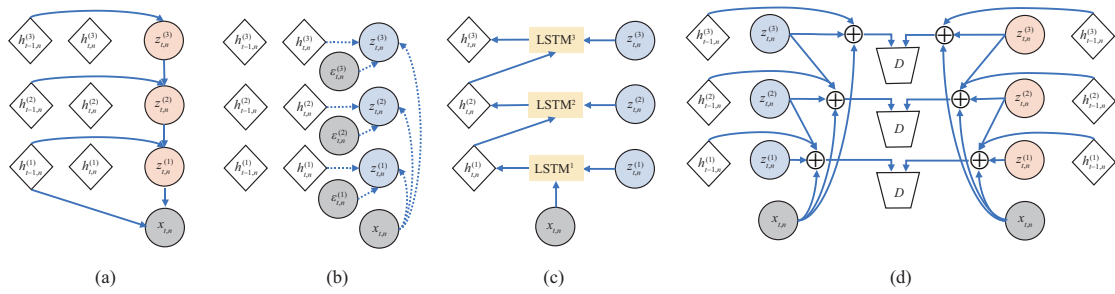


图 3 (网络版彩图) SaVRNN 的各步过程说明

Figure 3 (Color online) Graphical illustration of each operation of the SaVRNN. (a) Hierarchical generation; (b) inference; (c) stacked recurrent; (d) multilayer adversarial

## 4.2 完整框架

本文所提的 SaVRNN 模型的完整框架如图 2 所示, 它由一个具有多层循环结构的变分编码网络、一个卷积编码网络和一个反卷积解码网络构成. 根据前文介绍的异常检测的基本流程, 将预处理好的 MTS 数据输入到 SaVRNN 中进行离线的训练. 在开始与结束阶段使用卷积和反卷积网络首先对 MTS 数据中的局部结构特性进行捕捉, 并将得到的特征表示定义为  $\hat{x}_n$ . 之后, 基于多层的循环变分自编码模型建模 MTS 数据中的复杂时序相关性, 并进行数据的重构. 最后, 根据无监督异常检测的流程进行阈值选取和在线的检测与分析.

## 4.3 堆叠式对抗变分循环神经网络

与传统的浅层动态异常检测方法不同, 在本小节提出的深层次概率动态模型 SaVRNN, 它利用层次化先验和堆叠式循环结构描述多层次和长距离的时序相关性, 因而具有较强的模型表征能力. 模型各步过程如图 3 所示.

### 4.3.1 层次化生成过程

如图 3(a) 所示, SaVRNN 自上而下的概率生成过程可以表示为

$$\begin{aligned}
 z_{t,n}^{(L)} &\sim \left( \boldsymbol{\mu}_{t,n}^{(L)}, \text{diag} \left( \boldsymbol{\sigma}_{t,n}^{(L)} \right) \right), \boldsymbol{\mu}_{t,n}^{(L)} = f \left( \mathbf{W}_{h,\mu}^{(L)} \mathbf{h}_{t-1,n}^{(L)} \right), \boldsymbol{\sigma}_{t,n}^{(L)} = f \left( \mathbf{W}_{h,\sigma}^{(L)} \mathbf{h}_{t-1,n}^{(L)} \right), \\
 &\dots, \\
 z_{t,n}^{(l)} &\sim \mathcal{N} \left( \boldsymbol{\mu}_{t,n}^{(l)}, \text{diag} \left( \boldsymbol{\sigma}_{t,n}^{(l)} \right) \right), \boldsymbol{\mu}_{t,n}^{(l)} = f \left( \mathbf{W}_{z\mu}^{(l)} z_{t,n}^{(l+1)} + \mathbf{W}_{h\mu}^{(l)} \mathbf{h}_{t-1,n}^{(l)} \right), \boldsymbol{\sigma}_{t,n}^{(l)} = f \left( \mathbf{W}_{z\sigma}^{(l)} z_{t,n}^{(l+1)} + \mathbf{W}_{h\sigma}^{(l)} \mathbf{h}_{t-1,n}^{(l)} \right),
 \end{aligned}$$

...

$$\hat{\mathbf{x}}_{t,n} \sim \mathcal{N}(\boldsymbol{\mu}_{t,n}^x, \text{diag}(\boldsymbol{\sigma}_{t,n}^x)), \boldsymbol{\mu}_{t,n}^x = f(\mathbf{W}_{z\mu}^x \mathbf{z}_{t,n}^{(1)} + \mathbf{W}_{h\mu}^x \mathbf{h}_{t-1,n}), \boldsymbol{\sigma}_{t,n}^x = f(\mathbf{W}_{z\sigma}^x \mathbf{z}_{t,n}^{(1)} + \mathbf{W}_{h\sigma}^x \mathbf{h}_{t-1,n}). \quad (4)$$

接下来对该过程进行详细描述.  $\mathbf{z}_{t,n}^{(l)} \in \mathbb{R}^{d_l}$  表示输入  $\mathbf{x}_{t,n}$  第  $l$  层服从高斯分布的概率隐变量, 它的均值和方差分别为  $\boldsymbol{\mu}_{t,n}^{(l)}$  和  $\boldsymbol{\sigma}_{t,n}^{(l)}$ .  $\mathbf{h}_{t,n}^{(l)}$  表示第  $l$  层的确定性的循环隐状态, 我们使用全连接网络实现隐变量之间的非线性转换, 并且  $f(\cdot)$  表示确定性的非线性函数.  $\{\mathbf{W}_{z\mu}^{(l)}, \mathbf{W}_{z\sigma}^{(l)}\} \in \mathbb{R}^{d_{l+1} \times d_l}$  是  $l < L$  层的因子装载矩阵, 并且  $\{\mathbf{W}_{h\mu}^{(l)}, \mathbf{W}_{h\sigma}^{(l)}\} \in \mathbb{R}^{d_l \times d_l}$  是第  $l$  层连接循环隐状态  $\mathbf{h}_{t,n}^{(l)}$  和概率隐变量  $\mathbf{z}_{t,n}^{(l)}$  的转移矩阵. 可以看出, SaVRNN 将第  $l < L$  层概率隐变量  $\mathbf{z}_{t,n}^{(l)}$  的均值与方差分成了两部分: 一部分是  $l+1$  层概率隐变量和因子装载矩阵的乘积, 用来捕捉当前时刻层次化的结构特性. 另一部分则是时刻转移矩阵与循环隐状态的乘积, 为了捕捉当前层不同时刻特征之间的时序相关性. 在模型的最高层,  $\mathbf{z}_{t,n}^{(L)}$  只依赖于当前层的循环隐状态.

总结来说, SaVRNN 将堆叠循环神经网络 [35~37] (捕捉多层次和长距离的时序依赖), 层次化变分自编码 [38,39] (通过层次化先验增强表征能力) 和对抗训练策略 [19] (表征隐层的表征能力) 融合到一个定义清晰的贝叶斯框架下. 相似地, 模型的生成分布定义为均值和方差分别为  $\boldsymbol{\mu}_{t,n}^x$  和  $\boldsymbol{\sigma}_{t,n}^x$  的高斯分布, 并且该分布不仅依赖于概率隐状态  $\mathbf{z}_{t,n}^{(1)}$ , 还依赖于循环隐状态  $\mathbf{h}_{t-1,n}^{(1)}$ .  $\{\mathbf{W}_{z\mu}^x, \mathbf{W}_{z\sigma}^x\} \in \mathbb{R}^{d_1 \times V}$  和  $\{\mathbf{W}_{h\mu}^x, \mathbf{W}_{h\sigma}^x\} \in \mathbb{R}^{d_1 \times V}$  是我们生成模型中可学习的参数. 模型完整生成过程的图示说明列在图 3(a) 和 (b) 中. 需要说明的是, 在  $L = 1$  且忽视对抗训练的情况下, SaVRNN 退化为 VRNN. 我们将与 SaVRNN 相同结构, 但基于变分推理进行训练的模型称为 SVRNN.

### 4.3.2 堆叠式循环过程

受文献 [35~37] 的启发, 为了捕捉多种尺度的隐结构特征, 基于堆叠循环神经网络 (SRNN) 进行隐状态  $\mathbf{h}_{t,n}$  的更新. 详细地, 如图 3(c) 所示, 使用下式更新它的隐状态:

$$\mathbf{h}_{t,n}^{(l)} = \begin{cases} \text{LSTM}^l \left( \mathbf{h}_{t-1,n}^{(l)}, \text{FNN} \left[ \mathbf{x}_{t,n}, \mathbf{z}_{t,n}^{(l)} \right] \right), & \text{if } l = 1, \\ \text{LSTM}^l \left( \mathbf{h}_{t-1,n}^{(l)}, \text{FNN} \left[ \mathbf{h}_{t,n}^{(l-1)}, \mathbf{z}_{t,n}^{(l)} \right] \right), & \text{if } 1 < l \leq L. \end{cases} \quad (5)$$

这里  $\text{LSTM}^l$  表示第  $l$  层的长短时记忆网络 (LSTM) [40], 我们定义其参数为  $\boldsymbol{\theta}_l$ . 隐状态  $\mathbf{h}_{t,n}^{(l)}$  利用  $\mathbf{z}_{t,n}^{(l)}$ ,  $\mathbf{h}_{t,n}^{(l-1)}$  和  $\mathbf{h}_{t-1,n}^{(l)}$  进行更新.  $[\mathbf{x}_{t,n}, \mathbf{z}_{t,n}^{(l)}]$  表示将  $\mathbf{x}_{t,n}$  和  $\mathbf{z}_{t,n}^{(l)}$  进行拼接, 而 FNN 表示全连接网络 (fully connect neural network), 是为了将概率隐表示与动态隐状态变量进行结合, 将具有不确定与层次化结构特性的概率隐表示信息引入到堆叠循环神经网络中. 正因为我们的模型将堆叠循环神经网络与层次化概率生成模型相结合, 因此它不需要类似于文献 [35,36] 中专门的优化策略来保证层次化结构隐变量的捕捉.

### 4.3.3 向上 – 向下推理

按照变分自编码模型建模思想 [41,42], 为实现对隐变量后验的估计, 首先定义一个灵活的变分分布  $q(\mathbf{z}_{t,n} | \mathbf{x}_{t,n})$  来近似真实的后验分布  $p(\mathbf{z}_{t,n} | -)$ , 并将其分解为自下而上的结构:

$$q(\mathbf{z}_{t,n} | \mathbf{x}_{t,n}) = q(\mathbf{z}_{t,n}^{(1)} | \mathbf{x}_{t,n}) \prod_{l=1}^{L-1} q(\mathbf{z}_{t,n}^{(l+1)} | \mathbf{z}_{t,n}^{(l)}). \quad (6)$$

传统变分推理方法直接赋予变分分布  $q(\mathbf{z}_{t,n}^{(l)} | -)$  显式表示 (如高斯分布), 这往往不足以刻画真实的后验分布, 特别是后验分布形式非常复杂的分层概率生成模型. 我们引入隐式分布的变分隐变量  $q(\mathbf{z}_{t,n}^{(l)} | -)$

进行后验估计, 其具体定义如下:

$$q(\mathbf{z}_{t,n}^{(l)} | -) = q_\phi(-, \boldsymbol{\epsilon}_{t,n}^{(l)}), \quad (7)$$

其中,  $\boldsymbol{\epsilon}_{t,n}^{(l)} \sim \mathcal{N}(0, \mathbf{I})$ ,  $\phi$  表示推理模型的可学习参数. 为了避免分层贝叶斯结构导致的高层随机隐变量先验崩溃问题 [43, 44], 不仅引入一个确定性自下而上的信息传输, 将递归神经网络的隐状态与  $l = 1$  到  $l = L$  的多层隐变量联系起来, 而且引入了一个从输入  $\hat{\mathbf{x}}_{t,n}$  到概率隐表示的映射参数, 最终得到  $l$  层隐变量变分分布表示

$$\hat{q}(\mathbf{z}_{t,n}^{(l)} | -) = q_\phi(\hat{\mathbf{x}}_{t,n}, \mathbf{h}_{t-1}^{(l)}, \boldsymbol{\epsilon}_{t,n}^{(l)}). \quad (8)$$

然后, SaVRNN 将提取的隐特征与生成模型中自上向下路径的先验信息结合起来, 构造从  $l = 1$  到  $l = L$  的概率隐状态的变分后验.

$$q(\mathbf{z}_{t,n}^{(l)} | -) = q_\phi(\mathbf{x}_{t,n}, \mathbf{h}_{t-1}^{(l)}, \boldsymbol{\epsilon}_{t,n}^{(l)}, \mathbf{z}_{t,n}^{(l+1)}). \quad (9)$$

基于推理网络的结构, 如图 3(b), 通过对生成分布中自下而上的似然信息与自上而下的先验信息组合来估计 SaVRNN 中隐变量的后验概率. 推理网络的变分分布如下所示:

$$q(\mathbf{z}_{t,n} | \hat{\mathbf{x}}_{t,n}) = q_\phi(\mathbf{z}_{t,n}^{(L)} | \hat{\mathbf{x}}_{t,n}, \mathbf{h}_{t,n}^{(L)}) \prod_{l=1}^{L-1} q_\phi(\mathbf{z}_{t,n}^{(l)} | \mathbf{z}_{t,n}^{(l+1)}, \hat{\mathbf{x}}_{t,n}, \mathbf{h}_{t,n}^{(l)}). \quad (10)$$

可以看出, 一方面, SaVRNN 模型使高层的随机隐变量依赖于输入  $\hat{\mathbf{x}}_{t,n}$ , 从而保证隐变量能获取更多有效信息并且不易出现后验失效. 另一方面, 通过传递当前层的循环隐状态和高层隐变量的动态信息, 推理模型能够学习 SaVRNN 更加丰富的隐表示. 此外, 将隐式的变分分布赋予  $\mathbf{z}_{t,n}^{(l)}$  可以保证其具有足够的表征能力组合这些信息. 需要说明的是, 当 SaVRNN 的隐变量变分分布设置为协方差矩阵为对角阵的高斯分布时, 即  $q(\mathbf{z}_{t,n}^{(l)} | -) = \mathcal{N}(\hat{\boldsymbol{\mu}}, \text{diag}(\hat{\boldsymbol{\sigma}}))$ . 此时, 通过自编码变分推理方法便可实现模型的推理, 此时的模型称之为堆叠循环神经网络 (stacked RNN, SRNN), 它可以看作 SaVRNN 忽略对抗训练退化得到的模型.

#### 4.4 多层对抗

同普遍的 VAE 模型的训练策略一样, SaVRNN 的优化目标可以通过最大化 ELBO (evidence lower bound), 表达式如下:

$$\begin{aligned} \mathcal{L} = & \sum_{t=1}^T \sum_{n=1}^N \mathbb{E}_{q_\phi(\mathbf{z}_{t,n}^{(1)} | \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(1)})} \left[ \ln p_\theta(\mathbf{x}_{t,n} | \mathbf{z}_{t,n}^{(1)}, \mathbf{h}_{t-1,n}^{(1)}) \right] \\ & - \sum_{t=1}^T \sum_{n=1}^N \sum_{l=1}^L \text{KL} \left( q_\theta(\mathbf{z}_{t,n}^{(l)} | \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)}) p_\phi(\mathbf{z}_{t,n}^{(l)} | \mathbf{z}_{t,n}^{(l+1)}) \right), \end{aligned} \quad (11)$$

其中, 第 1 项表示对数似然的期望, 用于保证重构的性能, 第 2 项表示 KL 散度, 它约束变分分布  $q(\mathbf{z}_{t,n}^{(l)})$  与先验分布  $p(\mathbf{z}_{t,n}^{(l)})$  的相似度. 式 (11) 的优化问题可表示为

$$\begin{aligned} & \max_{\theta} \max_{\phi} \sum_{t=1}^T \sum_{n=1}^N \mathbb{E}_{p_D(\mathbf{x}_{t,n})} \mathbb{E}_{q_\phi(\mathbf{z}_{t,n}^{(1)} | \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(1)})} \ln p_\theta(\mathbf{x}_{t,n} | \mathbf{z}_{t,n}^{(1)}, \mathbf{h}_{t-1,n}^{(1)}) \\ & + \sum_{l=1}^L \mathbb{E}_{q_\phi(\mathbf{z}_{t,n}^{(l)} | \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)})} \left[ p(\mathbf{z}_{t,n}^{(l)} | \mathbf{z}_{t,n}^{(l+1)}) - q_\phi(\mathbf{z}_{t,n}^{(l)} | \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)}) \right]. \end{aligned} \quad (12)$$

受对抗变分贝叶斯方法 [19] 的启发, 为实现对隐状态的隐式变分分布的优化, 在每一层引入判别网络:

$$T_l \left( \mathbf{z}_{t,n}^{(l)}, \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)} \right), \quad l = 1, \dots, L, \quad \mathbf{z}_{t,n}^{(0)} = \mathbf{x}_{t,n}. \quad (13)$$

网络每层的优化目标如下:

$$\begin{aligned} \max_{T_l} & \sum_{t=1}^T \sum_{n=1}^N \mathbb{E}_{p_D(\mathbf{x}_{t,n})q_\phi(\mathbf{z}_{t,n}^{(l)})} \log \sigma \left( T_l \left( \mathbf{z}_{t,n}^{(l)}, \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)} \right) \right) \\ & + \mathbb{E}_{p_D(\mathbf{x}_{t,n})p(\mathbf{z}_{t,n}^{(l)}|\mathbf{z}_{t,n}^{(l+1)})} \log \left( 1 - \sigma \left( T_l \left( \mathbf{z}_{t,n}^{(l)}, \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)} \right) \right) \right). \end{aligned} \quad (14)$$

基于文献 [19] 中的推理过程, 每层的最优判别器为

$$T_l^* \left( \mathbf{z}_{t,n}^{(l)}, \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)} \right) = \log q_\phi \left( \mathbf{z}_{t,n}^{(l)} | \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)} \right) - \log p \left( \mathbf{z}_{t,n}^{(l)} | \mathbf{z}_{t,n}^{(l+1)} \right). \quad (15)$$

根据式 (9) 和分布的重参技巧, 优化目标可写为

$$\mathcal{L} = \max_{\theta} \max_{\phi} \sum_{t=1}^T \sum_{n=1}^N \left[ \mathbb{E}_{\epsilon_{t,n}^{(l)}} \left[ \ln p_{\theta} \left( \mathbf{x}_{t,n} | \mathbf{z}_{t,n}^{(1)} \right) \right] - \sum_{l=1}^L T_l^* \left( \mathbf{z}_{t,n}^{(l)}(\mathbf{x}_{t,n}, \epsilon_{t,n}^{(l)}), \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)} \right) \right]. \quad (16)$$

通过对每一层的对抗优化, 可以实现对隐式分布变分隐变量的推理. 图 3(d) 中给出本文提出的多层对抗过程的详细图解.

## 5 模型训练

分层对抗网络的优化非常困难, 受文献 [20] 启发, 引入一种基于条件传输 (CT) 的分层优化方法. 基于上述, 给出本文提出的 SaVRNN 模型训练过程.

### 5.1 基于条件传输的改进优化方法

如前文所述, 提出的分层对抗网络很难平衡生成器与判别器二者之间的更新, 并且在实验过程中也发现这一问题. 受文献 [20] 的启发, 引入 CT 框架用以修正我们的损失函数, 该框架在不要求分布共享相同支持度的条件下能够衡量两种离散分布之间的差异, 减轻模型的失衡问题. 基于此, 我们提出一种基于 CT 的 SaVRNN 优化方法. 具体来说, CT 中两种经验分布之间的传输损失表示为

$$\text{CT}(p, q) := \min_{\mathbf{T}^{p \rightarrow q}, \mathbf{T}^{q \rightarrow p}} \text{Tr} \left[ (\mathbf{T}^{p \rightarrow q})^T \mathbf{C} + (\mathbf{T}^{q \rightarrow p})^T \mathbf{C} \right], \quad (17)$$

其中,  $\text{Tr}(\cdot)$  表示矩阵的迹;  $\mathbf{C} \in \mathbb{R}^{n \times m}$  表示传输损失矩阵为  $C_{i,j} = c(\mathbf{x}_i, \mathbf{y}_j)$ ,  $i = 1 : n$ ,  $j = 1 : m$ .  $\mathbf{T}^{p \rightarrow q}$  和  $\mathbf{T}^{q \rightarrow p}$  分别表示正向和反向两种传输方式, 具体为

$$\mathbf{T}_{ij}^{p \rightarrow q} = u_i \frac{v_j e^{-d_\psi(\mathbf{y}_j, \mathbf{x}_i)}}{\sum_{j'=1}^m v_{j'} e^{-d_\psi(\mathbf{x}_i, \mathbf{y}_{j'})}}, \quad \mathbf{T}_{ij}^{q \rightarrow p} = v_j \frac{u_i e^{-d_\psi(\mathbf{y}_j, \mathbf{x}_i)}}{\sum_{i'=1}^n u_{i'} e^{-d_\psi(\mathbf{x}_{i'}, \mathbf{y}_j)}}. \quad (18)$$

这里定义  $\psi$  参数化的  $d_\psi(\mathbf{x}, \mathbf{y}) = d_\psi(\mathbf{y}, \mathbf{x})$  衡量两个向量之间的差异. 这里  $u_i$  和  $v_i$  分别为向量  $\mathbf{x}_i$  和  $\mathbf{y}_j$  对应分布的概率取值. 基于数据形式并借鉴已有工作 [21, 22], 假设它们  $u_i$  和  $v_i$  的取值服从均匀



分布, 因此  $u_i = 1/M$ ,  $v_i = 1/N$ . 显然, CT 是从双向角度考虑传输方式, 从而保证多层生成和对抗网络的收敛性. 具体来说, 为了借助 CT 优化本文提出的 SaVRNN 模型, 定义  $Q$  和  $P$  的分布为

$$\begin{aligned} \boldsymbol{\epsilon}_{t,n,j}^{(l)} &\sim \mathcal{N}(0, \mathbf{I}), \quad \mathbf{z}_{t,n}^{(l)} \sim p(\mathbf{z}_{t,n}^{(l)}), \\ \mathbf{x}_{t,n,i}^{(l)} &= \tilde{\mathbf{T}}_l(\tilde{\mathbf{z}}_{t,n,j}^{(l)}(\mathbf{x}, \boldsymbol{\epsilon}_{t,n,j}^{(l)}), \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)}), \quad \mathbf{y}_{t,n,j} = \tilde{\mathbf{T}}_l(\mathbf{z}_{t,n}^{(l)}, \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)}), \\ \mathbf{v} = P(\mathbf{y}_{t,n}^{(l)}) &= \frac{1}{M} \sum_{j=1}^M \delta(\mathbf{y}_{t,n}^{(l)} - \mathbf{y}_{t,n,j}^{(l)}), \quad \mathbf{u} = Q(\mathbf{x}_{t,n}^{(l)}) = \frac{1}{N} \sum_{i=1}^N \delta(\mathbf{x}_{t,n}^{(l)} - \mathbf{x}_{t,n,i}^{(l)}). \end{aligned} \quad (19)$$

由式 (18), 定义向前向后传输表达式为

$$\mathbf{T}_{ij}^{p \rightarrow q} = u_i \frac{v_j e^{-d_\psi(\mathbf{y}_j, \mathbf{x}_i)}}{\sum_{j'=1}^m v_{j'} e^{-d_\psi(\mathbf{x}_i, \mathbf{y}_{j'})}}, \quad \mathbf{T}_{ij}^{p \leftarrow q} = v_j \frac{u_i e^{-d_\psi(\mathbf{y}_j, \mathbf{x}_i)}}{\sum_{i'=1}^n u_{i'} e^{-d_\psi(\mathbf{x}_{i'}, \mathbf{y}_j)}}. \quad (20)$$

具体来说, 令  $c(\cdot)$  和  $d(\cdot)$  的表达式为

$$c(\mathbf{x}_i, \mathbf{y}_j) = e^{-\mathbf{x}_i^\top \mathbf{y}_j}, \quad d_\psi(\mathbf{x}_i, \mathbf{y}_j) = -\mathbf{x}_i^\top \mathbf{y}_j. \quad (21)$$

进而, 可以得到基于 CT 的优化目标为

$$\begin{aligned} \mathcal{L}_{ct} &= \sum_{n=1}^N \sum_{t=1}^T \left[ \sum_{i=1}^N e^{-\mathbf{x}_i^\top \mathbf{y}_j} \frac{1}{N} \frac{e^{\mathbf{x}_i^\top \mathbf{y}_j}}{\sum_{ii=1}^N e^{\mathbf{x}_{ii}^\top \mathbf{y}_j}} + \sum_{j=1}^M e^{-\mathbf{x}_i^\top \mathbf{y}_j} \frac{1}{M} \frac{e^{\mathbf{x}_i^\top \mathbf{y}_j}}{\sum_{jj=1}^M e^{\mathbf{x}_i^\top \mathbf{y}_{jj}}} \right] \\ &= \sum_{n=1}^N \sum_{t=1}^T \left[ \sum_{i=1}^N \frac{1}{N \sum_{ii=1}^N e^{\mathbf{x}_{ii}^\top \mathbf{y}_j}} + \sum_{j=1}^M \frac{1}{M \sum_{jj=1}^M e^{\mathbf{x}_i^\top \mathbf{y}_{jj}}} \right]. \end{aligned} \quad (22)$$

结合式 (15), 判别模块的损失函数可以表示为

$$L_d(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^N \sum_{t=1}^T \left[ \sum_{i=1}^N \frac{\rho}{N \sum_{ii=1}^N e^{\mathbf{x}_{ii}^\top \mathbf{y}_j}} + \sum_{j=1}^M \frac{1-\rho}{M \sum_{jj=1}^M e^{\mathbf{x}_i^\top \mathbf{y}_{jj}}} \right]. \quad (23)$$

更多关于条件传输原理的讲解可以参考文献 [20]. 我们利用式 (23) 作为损失函数去优化每一层判别器网络式 (13).

## 5.2 训练过程

与 VAE 模型类似, SaVRNN 中推理与生成网络参数的优化目标可以通过最大化 ELBO, 其表达式为

$$\mathcal{L} = \max_{\theta} \max_{\phi} \sum_{t=1}^T \sum_{n=1}^N \left[ \mathbb{E}_{\boldsymbol{\epsilon}_{t,n}^{(l)}} \left[ \ln p_{\theta}(\mathbf{x}_{t,n} | \mathbf{z}_{t,n}^{(1)}) \right] - \beta^{(l)} \sum_{l=1}^L \mathbf{T}_l^* \left( \mathbf{z}_{t,n}^{(l)}(\mathbf{x}, \boldsymbol{\epsilon}_{t,n}^{(l)}), \mathbf{z}_{t,n}^{(l+1)}, \mathbf{x}_{t,n}, \mathbf{h}_{t-1,n}^{(l)} \right) \right], \quad (24)$$

其中, 第 1 项是对数似然的期望, 用于保证模型的重构性能. 第 2 项是對抗項, 用于约束变分分布  $q(\mathbf{z}_{t,n}^{(l)})$  与先验分布  $p(\mathbf{z}_{t,n}^{(l)})$  相似度. 同时, 通过损失函数 (23) 优化对抗网络中分类器 (13) 的参数. 如算法 1 描述, SaVRNN 的生成、推理和递归模块的参数分别定义为  $\{\boldsymbol{\Omega}^{(l)}\}_{l=1}^L$ ,  $\{\boldsymbol{\Psi}^{(l)}\}_{l=1}^L$ ,  $\boldsymbol{\theta}$ , 采用随机梯度下降法 (stochastic gradient descent, SGD) 对其共同更新. 即使采用 CT 修正损失, SaVRNN 作为一种深度分层动态概率生成模型, 熟知的“后验失效”仍是一个巨大挑战 [45, 46]. 为了解决这个问题, 我们引入一种方法, 称之为“warm-up”, 用于稳定训练 [17]. “warm-up”首先只利用重构误差训练参数, 然后在第 1 个训练周期内, 逐步加入对抗损失, 其系数  $\beta$  逐渐由 0 增加到 1, 从而有效缓解“后验失效”问题.

**算法 1** 面向 SaVRNN 模型的向上 - 向下对抗变分推理

**输入:** 预处理后的 MTS 训练集  $\mathcal{D}(\mathbf{x}_{1:T})$ ;

**输出:** SaVRNN 推理模型参数:

$$\Omega^{(l)} = \{C_{x\mu}^{(l)}, C_{x\sigma}^{(l)}, \bar{C}_{h\mu}^{(l)}, \bar{C}_{h\sigma}^{(l)}, \bar{C}_{h\mu}^{(l)}, \bar{C}_{h\sigma}^{(l)}\}, l = 1, \dots, L;$$

SaVRNN 生成模型参数:

$$\Psi^{(l)} = \{\bar{W}_{h\mu}^{(l)}, \bar{W}_{h\sigma}^{(l)}, \bar{W}_{h\mu}^{(l)}, \bar{W}_{h\sigma}^{(l)}, W_{z\mu}^x, \bar{W}_{h\mu}^x, \bar{W}_{h\sigma}^x, W_{z\sigma}^x, \bar{W}_{h\mu}^x, \bar{W}_{h\sigma}^x\};$$

循环神经网络参数  $\theta_l$ ; 卷积与反卷积网络参数  $D$ ; 对抗分类网络参数  $d$ ;

- 1: 设定迷你块大小为  $M$ , 卷积核的数目为  $K$ , 设定超参数;
- 2: 初始化  $\{\Omega^{(l)}\}_{l=1}^L$ ,  $\{\Psi^{(l)}\}_{l=1}^L$ ,  $\theta$  和  $D$ ;

**主迭代:**

- 3: 随机选取一个由  $T$  个子序列组成小批量为  $M$  的 MTS, 构成迷你块  $\{\mathbf{x}_{1:T,i}\}_{i=1}^M$ ;
- 4: 从均匀分布中采样随机噪声  $\{\epsilon_{t,n}^{(l)}\}_{t=1,n=1,l=1}^{T,M,L}$ , 进而采样隐状态  $\{\mathbf{z}_{t,n}^{(l)}\}_{t=1,n=1,l=1}^{T,M,L}$ ;
- 5: **for**  $l = 1, 2, \dots, L$  **do**
- 6: 基于式 (24) 计算  $\nabla_{\Omega^{(l)}, \Psi^{(l)}, D, \theta_l} L(\Omega^{(l)}, \Psi^{(l)}; \mathbf{X}, \epsilon_{t,n}^{(l)})$ , 更新编码器参数  $\Omega^{(l)}$ , 解码器参数  $\Psi^{(l)}$  和循环网络参数  $\theta_l$ ;
- 7: 通过损失函数 (23) 优化对抗网络中分类器 (13) 的参数  $d$ ;
- 8: **end for**
- 9: 更新卷积与反卷积网络参数  $D$ ;

**输出:** 更新得到的编码器、解码器、循环网络、判别网络和卷积反卷积网络参数.

## 6 异常检测

同目前普遍采用的无监督异常检测策略一样, 训练 SaVRNN 用于学习 MTS 的正常模式, 然后利用  $x_{t,n}$  的重构概率作为异常评分来判断观测变量是否异常, 异常评分计算方式如下:

$$\mathcal{S}_{t,n} = \log p(\mathbf{x}_{t,n} | \mathbf{z}_{t,n}), \quad (25)$$

其中, 得分越高表示重构性能越好, 表明当前的观察越遵循正常模式. 然后基于异常分数进行阈值的指定, 如果  $\mathcal{S}_{t,n}$  低于此阈值, 则对应的  $\mathbf{x}_{t,n}$  被标记为异常. 在无监督异常检测中, 得到更好的正常数据模式的表示, 进而获得更好的拟合性能是至关重要的. 如前文所述, SaVRNN 具有层次化结构, 能够在不同层中利用其长距离的动态模式获得多层次的表征形式. 为了更好地利用多层次的表征并提升模型对输入的鲁棒性, 将异常评分重新定义为多层似然的联合条件概率, 可以表示为

$$\begin{aligned} \hat{\mathcal{S}}_{t,n} &= \frac{1}{L} \log p(\mathbf{x}_{t,n}, \mathbf{z}_{t,n}^{(1)}, \dots, \mathbf{z}_{t,n}^{(L-1)} | \mathbf{z}_{t,n}^{(L)}) \\ &= \frac{1}{L} \left( \log p(\mathbf{x}_{t,n} | \mathbf{z}_{t,n}^{(1)}) + \sum_{l=1}^{L-1} \log p(\mathbf{z}_{t,n}^{(l)} | \mathbf{z}_{t,n}^{(l+1)}) \right). \end{aligned} \quad (26)$$

它可以看作利用各层的重构表现共同进行异常的检测. 同文献 [1,47] 一样, 本文采用 POT (peaks-over-threshold) 方法帮助选择检测阈值. 基于训练好的 SaVRNN 模型, 将序列  $\mathbf{x}_{t-W:t}$  作为模型的输入, 重构  $\mathbf{x}_t$ , 然后根据式 (26) 计算异常评分, 最终实现对 MTS 数据的在线异常检测.

## 7 实验

本节进行了包括定量对比和定性分析的实验来验证 SaVRNN 的有效性.

表 1 数据的基本统计信息  
Table 1 Basic statistics of datasets

Statistics	DND	SMD	MSL
Datasize	12	28	27
Dimensions	36	38	55
Granularity (s)	60	60	60
Training set size	344843	708405	58317
Testing set size	344844	708420	73729
Anomaly ratio (%)	3.44	4.16	10.72

## 7.1 实验设定

**基准方法.** 将 SaVRNN 和一些目前优秀的 MTS 在线异常检测方法比较: (1) 非动态方法, 包括: 变分自编码模型 (VAE, 一种基础的概率生成模型); 阶梯变分自编码 (ladder VAE, 一种多层 VAE 结构的模型). (2) 动态方法, 包括: LSTM-NTD<sup>[13]</sup> (一种确定性的循环自编码模型); 变分神经网络<sup>[29]</sup> (VRNN, 一种可以建模高维时间序列的概率动态模型); OmniAnomaly<sup>[1]</sup> (一种循环网络结构的概率生成模型); SDFVAE<sup>[4]</sup> (一种基于静态与动态因子分解的 VAE 框架模型); DVGCRN<sup>[33]</sup> (将图卷积结构引入到变分神经网络, 是目前针对 MTS 异常检测任务性能最优的基于循环网络结构的模型).

**数据集.** 基于 3 个广泛使用的实测多维时间序列数据集进行实验, 包括: 一个多维关键表现指示 (key performance indicators, KPIs) 数据, 来自于分布式网络数据 (distributed network dataset, DND), 和两个公开数据集 SMD (Server Machine Dataset) 和 MSL (Mars Science Laboratory), 这 3 个数据集分别由文献 [1, 13] 公开. 数据集的详细信息见表 1. 详细地, 表 1 中的 Datasize 代表一个数据集中的多维时间序列的数目  $N$ , Dimension 代表一个多维时间序列中的通道数目  $V$ , Training set size 和 Testing set size 分别代表训练数据和测试数据中多维时间序列的时刻长度  $T$ , 而 Granularity 代表的是两个相邻时刻之间的采样时间. Anomaly ratio 代表数据中的异常时刻的百分比数目. DND KPI 数据集来自中国一个著名的 ISP 运营的 DND, 数据集包含 12 个网站, 分别由 36 个 KPI 监测. 这些网站的服务类型各不相同, 包括视频点播 (VoD), 直播视频等. 对于每个网站, KPIs 每 60 s 收集一次, 持续大约一个半月. 在本文实验中, 每个网站前半部分 KPI 数据用于训练, 后半部分用于测试. 值得注意的是, DND 测试集中被标注为异常的数据已经被研究人员证实. 对于公共数据集, 服务器数据集 (SMD) 是实际采集的数据集<sup>[1]</sup>, 包含 28 台服务器的数据, 每台机器分别由 38 个 KPI 监测. 另外, 对于每台服务器, KPI 的持续时间约为 5 周. 火星科学实验室 (MSL) 数据集也是 NASA<sup>[13]</sup> 提供的真实公开并由专家标记的数据集, 包含 27 个实体数据, 每个实体包含 55 个检测指标. MSL 由航天器采集, 其第一维是遥测信道的值, 其余维是编码为 0 或 1 的命令信息.

**超参数.** 实验中, 所有的基准方法均使用对应发表文章中公开的性能或者使用官方公开的最优参数设置与代码进行复现实现. 此外, 我们基于 Pytorch 实现了 SaVRNN. 卷积编码器和反卷积解码器 (如 DND 数据集) 均包含 3 个一维卷积层 (卷积核和步长分别为 (3,3), (2,2), (2,2)). 采用 Adam 优化器, 学习率为 0.0002, 迷你块大小设置为 256, 数据整体更新次数设置为 50. 另外, 在前 10 个训练阶段, 设定 ELBO 方程中的温度系数从 0 增加到 1. 根据经验, 将 POT 使用的与初始阈值相关的概率设定为 0.004.

**硬件平台.** 所有实验均在 Intel(R)Xeon(R)CPU E5 - 2667v3 @ 3.20 GHz 上进行, 采用两个 NVIDIA RTX8000 加速, 每块显卡均有 48 GB VRAM.

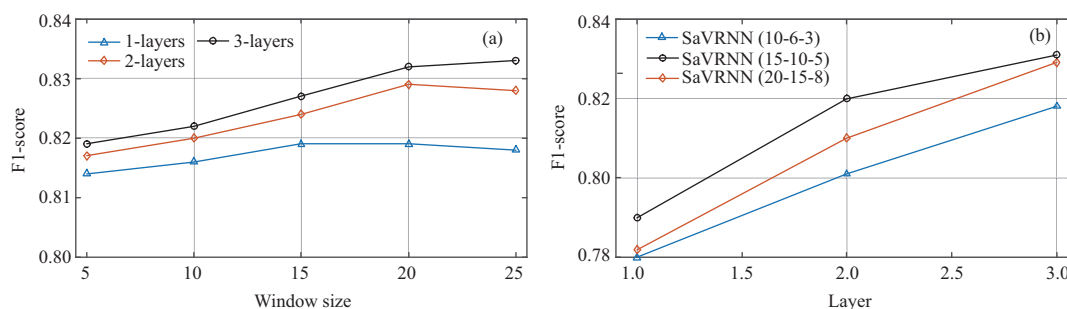


图 4 (网络版彩图) SaVRNN 在 SMD 数据集上的检测性能随窗长 (a) 和网络大小 (b) 的变化

Figure 4 (Color online) Detection accuracy of SaVRNN on SMD as a function of window size (a) and network size (b)

**评价指标.** 同文献 [1, 2, 4] 一样, 使用 Precision, Recall 和 F1-score 作为评价指标来衡量不同方法的性能. 特别是 F1-score, 被认为是一个综合指标, 因为它平衡了精准率和召回率.

## 7.2 定量对比

**参数敏感度.** 本文测试异常检测性能对于模型参数的敏感性, 包括模型层数  $L$ , 隐变量的维度  $\{d_i\}_{i=1}^L$  和窗长  $T$ .

使用上文所述的参数设置, 并将窗长设置为 20, 我们首先研究网络结构  $L$  和  $\{d_i\}_{i=1}^L$  对异常检测性能的影响, 结果如图 4 所示. 可以看出, 在给定第 1 层宽度条件下, 增加网络的深度能够明显提升异常检测精度, 这表明层次结构能够有效增强网络的表征能力和拟合能力. 此外, 在给定隐层深度的条件下, 随着隐层宽度的增加, SaVRNN 的异常检测性能并不会一直变好, 在某些点宽度的增加并没有帮助性能的提升, 例如图中 15-10-5 结构的模型是比 20-15-8 结构具有更优性能的. 我们认为造成这一现象的原因在于较大的隐状态维度带来了更高的计算成本和更多的训练样本需求.

在设定网络结构为 [15, 10, 5] 后, 我们测试窗长对性能的影响, 它决定了具有相关性的时刻范围. 图 4 给出  $T$  在 5~25 范围内模型检测精度的变化曲线. 显然, 随着窗长  $T$  长度的增加, 不同层数的 SaVRNN 检测精度呈现共同趋势, 即先增长后保持稳定, 这可以归结于模型只能捕捉有限长度的时刻间的相关性. 对于具有更深层的 SaVRNN 模型能够捕捉更长范围时刻间的相关性, 因而随着  $T$  增加, 其性能提升越大.

**异常检测性能.** 将本文提出的模型与之前介绍的所有基线模型的检测性能比较, 并将 5 次独立运行的结果的平均准确率列于表 2 中, 可以看出, 相比于非动态模型 VAE, ladder VAE 等, 具有动态结构的模型就有更好的 F1 分数性能, 说明时序特征捕捉在多维时间序列建模中的重要性. 此外, 多层结构模型比相对应的单层结构模型性能更优秀, 说明了多层结构带来的表征能力的提高. 本文提出的多层动态变分网络相比之前方法有着更优秀的性能, 这归功于层次概率生成和堆叠循环结构. 最后, 对于在线异常检测任务, 具有 3 个隐层的 SaVRNN 在所有检测方法中达到最好的检测性能, 进一步说明了多层对抗训练可以帮助模型捕捉表征能力更强的隐表示特征. 需要说明的是, VRNN 是 SaVRNN 没有多层结构和对抗训练的特殊形式, SVRNN 是 SaVRNN 没有对抗训练的特殊形式, SaVRNN 的网络结构为 20 时是其单层结构形式, 而 LSTM-NTD 是没有变分结构的动态网络, 这几个结果的对比充分说明了本文所提方法的多层结构、变分自编码框架, 以及对抗训练优化方式的优势. ‘-M’ 表示用式 (26) 中改进后的异常分数计算方法进行检测, 结果证明本文异常分数计算改进方法的有效性. 此外, 基于本文所提的多层对抗网络建模方法, 进一步对 DVGCRN 网络进行改进, 将其概率隐状态设

表 2 不同模型在 3 个数据集上检测性能表现  
 Table 2 Detection performance of different methods on three datasets

Methods	LSTM size	Latent size	DND			SMD			MSL		
			Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
VAE	-	15	0.912	0.635	0.748	0.989	0.685	0.809	0.781	0.813	0.797
LSTM-NTD	20	-	0.549	0.614	0.579	0.568	0.644	0.604	0.593	0.537	0.564
VRNN	20	15	0.802	0.777	0.789	0.970	0.795	0.874	0.884	0.902	0.893
ladder VAE	-	15-10-5	0.913	0.934	0.765	0.990	0.753	0.854	0.842	0.823	0.832
OmniAnomaly	20	15	0.919	0.723	0.809	0.819	0.968	0.887	0.887	0.912	0.899
SDFVAE	20	15	0.964	0.711	0.818	0.981	0.850	0.911	0.853	0.894	0.873
SVRNN	20-15	15-10	0.902	0.751	0.819	0.959	0.835	0.893	0.881	0.924	0.902
	20-15-10	15-10-5	0.929	0.754	0.832	0.960	0.850	0.902	0.890	0.922	0.906
SVRNN-M	20-15-10	15-10-5	0.930	0.756	<b>0.834</b>	0.949	0.878	<b>0.912</b>	0.886	0.930	<b>0.909</b>
SaVRNN	20	15	0.940	0.690	0.796	0.795	0.989	0.881	0.929	0.874	0.901
	20-15	15-10	0.823	0.829	0.825	0.957	0.852	0.901	0.939	0.880	0.908
	20-15-10	15-10-5	0.920	0.781	0.841	0.960	0.861	0.908	0.940	0.879	0.909
SaVRNN-M	20-15-10	15-10-5	0.911	0.782	<b>0.842</b>	0.950	0.879	<b>0.913</b>	0.941	0.888	<b>0.914</b>
DVGCRN	20	15	0.822	0.831	0.826	0.959	0.850	0.901	0.889	0.920	0.904
	20-15	15-10	0.929	0.754	0.832	0.960	0.850	0.902	0.890	0.922	0.906
	20-15-10	15-10-5	0.930	0.756	0.834	0.950	0.876	0.912	0.886	0.930	0.908
DVGCRN-M	20-15-10	15-10-5	0.929	0.782	0.849	0.950	0.883	0.915	0.888	0.941	0.914
DaVGRN-M	20-15-10	15-10-5	0.935	0.803	<b>0.864</b>	0.956	0.894	<b>0.923</b>	0.902	0.949	<b>0.925</b>

置为隐性分布变量, 并基于本文所提的优化方法进行模型训练, 将该模型称为深度对抗图卷积循环网络 (deep adversarial variational graph convolutional recurrent network, DaVGRN), 并将结果列在表 2 中. 可以看出, 引入隐式分布的隐变量和对抗训练之后, 模型的隐层具有了更强的表征能力, 进而使得 DaVGRN 的性能在 DVGCRN 的基础上实现进一步的提升. 此外, 也可以发现, 相比于 SVRNN, 对于 DVGCRN 这种特征挖掘与表征能力更强的深度概率模型, 隐式分布的隐变量和对抗训练可以带来更加显著的性能提升, 原因在于对于能力更强的生成模型, 简单分布的隐表示更加无法满足其表征能力要求, 进一步说明了本文所提深度对抗变分模型的优势.

### 7.3 消融实验

这里进一步通过消融实验来证明本文所提方法中各个结构的重要性, 主要包括循环网络结构、隐层不确定性引入、层次化生成结构, 以及对抗训练方式. 基于此目标, 我们将 VAE 作为基础模型, 然后对比 SaVRNN w/o probability (深度循环神经网络模型)、SaVRNN w/o stacked (将模型深度设为 1)、SaVRNN w/o recurrent (将时刻数据设为 1)、SaVRNN w/o adversarial (不使用对抗训练方法得到隐表示), 以及 SaVRNN. 我们将实验测试结果展示在图 5. 可以看到, 本文所提方法中的每一个结构和设定的引入都可以带来性能的提升, 说明了本文方法中的各个结构的高效性. 此外, SaVRNN w/o probability 的性能相比于基础模型 VAE 更差, 充分说明了在建模复杂的多维时间序列时, 不确定性引入的重要性.

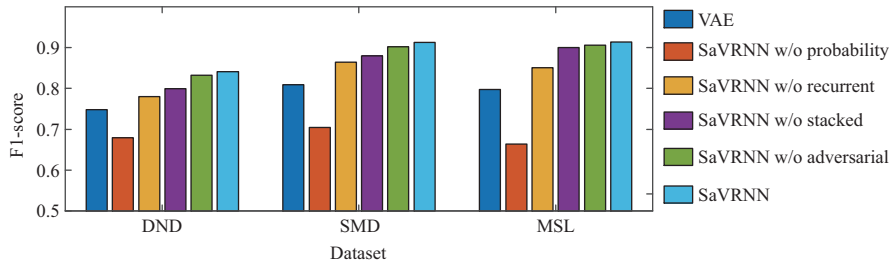


图 5 (网络版彩图) SaVRNN 在 3 个数据集上的消融实验  
 Figure 5 (Color online) Ablation study of SaVRNN on three datasets

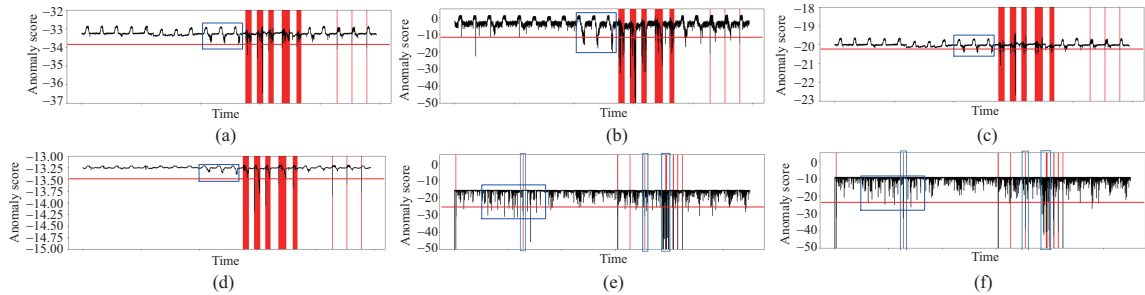


图 6 (网络版彩图) 案例分析: 基于 SaVRNN 在 SMD 的异常分数 (a)~(d), SVRNN 与 SaVRNN 在 DND 数据集上的异常分数 (e) 和 (f). 红线表示检测阈值

Figure 6 (Color online) Case study of anomaly score by (a) SaVRNN-3, (b) SaVRNN-1, (c) SaVRNN- $p(z_{t,n}^{(1)}|z_{t,n}^{(2)})$ , and (d) SaVRNN- $p(z_{t,n}^{(2)}|z_{t,n}^{(3)})$  on SMD machine-1-1 datasets; case study of anomaly score by (e) SVRNN and (f) SaVRNN on DND Website-1-10 datasets.

### 7.4 定性分析

为了进一步说明分层结构和对抗训练的有效性, 本文还做了一些定性分析.

**异常得分.** 为了验证模型中分层概率框架和堆叠递归结构以及对抗训练方式的有效性, 对不同层 SaVRNN 模型的异常分数进行了可视化研究, 结果如图 6 所示. 具体来说, 图 6(a) 展示了 3 层 SaVRNN 网络原始异常分数, 图 6(b) 展示了单层模型的异常分数. 明显地, 多层模型异常分数值更加稳定, 而在异常区域异常数值呈现很强的尖峰, 这进一步证明 SaVRNN 多层结构学习复杂 MTS 数据中正常模式的高效性. 此外, 在图 6(c) 和 (d) 中, 还对隐状态的对数似然进行可视化分析, 有趣的是, 它们在正常数据的时间步中是稳定的, 在异常数据的时间步中波动剧烈, 这使得修正后的异常分数保持平稳, 因而说明多层结构和我们提出修正检测方法的优越性. 并且, 我们还对 SVRNN 和 SaVRNN 的异常评分进行可视化, 进而说明对抗变分训练方式模型在捕捉异常模式方面优于变分训练方式. 从图 6 (e) 和 (f) 所示的结果看, SaVRNN 在异常区域具有更多的尖峰, 如图中蓝色线框所框定的时间步, 这足以说明 SaVRNN 能够有效检测异常模式.

**多层概率隐空间的可视化.** 为进一步说明多层结构能够有效捕捉具有层次化特性的特征, 构建大小为 12-6-3 的三隐层 SaVRNN, 并在数据集 DND 对其概率隐变量进行可视化分析, 结果如图 7 所示. 显然, 低层的隐表示比高层的隐表示变化更加剧烈, 体现了相邻时间步中细微的时序相关性. 然而, 来自高层的隐表示变化更加平滑, 使其能够捕捉跨时间步中的时序相关性. 而且, 无论是底层还是高层, 在异常的时间步中均存在特征波动, 这表明各层特征对异常检测的有效性, 也证明了本文多层建模及

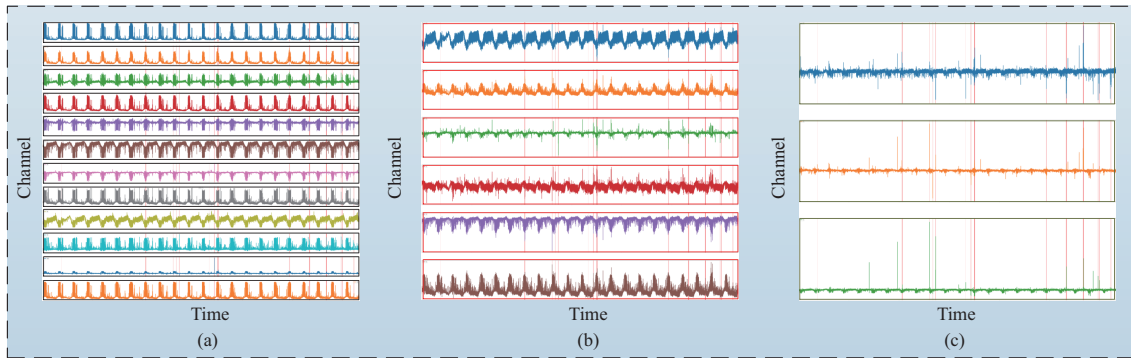


图 7 (网络版彩图) 案例分析: SaVRNN 在 DND 数据集网站 Weibsite-1-1 所学隐层特征可视化

Figure 7 (Color online) Case study of the sampled probabilistic latent variables of SaVRNN DND Website-1-1 datasets. (a) Layer1; (b) layer2; (c) layer3

表 3 本文所提模型的时间效率

Table 3 Training and testing times of the proposed methods

Time	Model	DND	SMD	MSL
Training times per epoch (min)	SVRNN	0.52	0.48	0.03
	SaVRNN	0.97	0.83	0.05
Testing times per sample (s)	SVRNN	0.068	0.069	0.084
	SaVRNN	0.126	0.129	0.155

利用多层似然进行异常检测的合理性.

### 7.5 时间效率

与先前文献 [4] 的工作类似, 我们研究了网络规模为 15-10-5 的 SVRNN 和 SaVRNN 网络训练和测试的时间效率, 结果如表 3 所示. 可以看出, 所提模型均能够在 0.1 s 内实现对单个样本的异常检测, 相比之下, 数据采集时间间隔为 60 s, 这说明 SVRNN 与基于对抗训练的 SaVRNN 可以用于在线检测.

## 8 结论

本文提出堆叠式对抗变分循环神经网络 (SaVRNN), 一种层次概率动态模型, 一方面引入层次化动态生成模型捕捉多维时间序列的层次结构特征和长距离时间相关性, 并借助多层似然改进检测方法实现异常检测, 另一方面引入隐式分布的隐变量与对抗训练的模型优化方式, 来保证隐层的表征能力. 为实现对所提模型精确的贝叶斯推理, 提出一种有效的向上 - 向下的对抗自编码推理方式. 此外, 为了保证模型的收敛, 引入条件最优传输 (CT) 方法进行损失函数的优化. 最后, 基于一个实测和两个主流公开数据集, 设计了充分的定性定量实验说明本文提出的模型在多维时间序列异常检测任务上具有高效性和优越性.

### 参考文献

- 1 Su Y, Zhao Y J, Niu C H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent

- neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery, 2019. 2828–2837
- 2 Su Y, Zhao Y, Sun M, et al. Detecting outlier machine instances through gaussian mixture variational autoencoder with one dimensional CNN. *IEEE Trans Comput*, 2022, 71: 892–905
  - 3 Sun M, Su Y, Zhang S L, et al. CTF: anomaly detection in high-dimensional time series with coarse-to-fine model transfer. In: Proceedings of IEEE Conference on Computer Communications, 2021. 1–10
  - 4 Dai L, Lin T, Liu C, et al. SDFVAE: static and dynamic factorized VAE for anomaly detection of multivariate CDN KPIs. In: Proceedings of the Web Conference, 2021. 3076–3086
  - 5 Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. *ACM Comput Surv*, 2009, 41: 1–58
  - 6 Chen W C, Xu H W, Li Z Y, et al. Unsupervised anomaly detection for intricate KPIs via adversarial training of VAE. In: Proceedings of IEEE Conference on Computer Communications, 2019. 1891–1899
  - 7 Chen Y Y, Mahajan R, Sridharan B, et al. A provider-side view of web search response time. In: Proceedings of ACM SIGCOMM Computer Communication Review, 2013
  - 8 Liu D P, Zhao Y J, Xu H W, et al. Opprentice: towards practical and automatic anomaly detection through machine learning. In: Proceedings of ACM Internet Measurement Conference, 2015. 211–224
  - 9 Xu H W, Chen W X, Zhao N W, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. In: Proceedings of the Web Conference, 2018. 187–196
  - 10 Shon T, Moon J. A hybrid machine learning approach to network anomaly detection. *Inf Sci*, 2007, 177: 3799–3821
  - 11 Yamada M, Kimura A, Naya F, et al. Change-point detection with feature selection in high-dimensional time-series data. In: Proceedings of International Joint Conference on Artificial Intelligence, 2013. 1827–1833
  - 12 Audibert J L, Michiardi P, Guyard F, et al. USAD: unsupervised anomaly detection on multivariate time series. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery, 2020. 3395–3404
  - 13 Hundman K, Constantinou V, Laporte C, et al. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery, 2018. 387–395
  - 14 Li D, Chen D C, Jin B H, et al. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks. In: Proceedings of International Conference on Artificial Neural Networks, 2019. 703–716
  - 15 Malhotra P, Ramakrishnan A, Anand G, et al. LSTM-based encoder-decoder for multi-sensor anomaly detection. In: Proceedings of International Conference on Machine Learning, 2016
  - 16 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. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence, 2019. 1409–1416
  - 17 Duan Z B, Wang D S, Chen B, et al. Sawtooth factorial topic embeddings guided gamma belief network. In: Proceedings of International Conference on Machine Learning, 2021. 2903–2913
  - 18 Sønderby C K, Raiko T, Maaløe L, et al. Ladder variational autoencoders. In: Proceedings of International Conference on Neural Information Processing Systems, 2017. 3738–3746
  - 19 Mescheder L, Nowozin S, Geiger A. Adversarial variational Bayes: unifying variational autoencoders and generative adversarial networks. In: Proceedings of International Conference on Machine Learning, 2017. 2391–2400
  - 20 Zheng H J, Zhou M Y. Exploiting chain rule and Bayes theorem to compare probability distributions. In: Proceedings of International Conference on Neural Information Processing Systems, 2021. 14993–15006
  - 21 Zheng H, Zhou M. Comparing probability distributions with conditional transport. 2020. ArXiv:2012.14100
  - 22 Korawat T, Fan X, Zheng H, et al. A prototype-oriented framework for unsupervised domain adaptation. In: Proceedings of International Conference on Neural Information Processing Systems, 2021. 17194–17208
  - 23 Zong B, Song Q, Min M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: Proceedings of International Conference on Learning Representations, 2018
  - 24 Dai L, Chen W, Liu Y, et al. Switching Gaussian mixture variational RNN for anomaly detection of diverse CDN websites. In: Proceedings of IEEE Conference on Computer Communications, 2022. 300–309
  - 25 Linderman S W, Miller A C, Adams R. Recurrent switching linear dynamical systems. In: Proceedings of International Conference on Artificial Intelligence and Statistics, 2016
  - 26 Ehmck P, Peters J, Patrick S. Switching linear dynamics for variational Bayes filtering. In: Proceedings of International Conference on Machine Learning, 2019. 553–562



- 27 Chen W, Chen B, Liu Y, et al. Switching poisson gamma dynamical systems. In: Proceedings of International Joint Conference on Artificial Intelligence, 2020. 2029–2036
- 28 Chen W, Chen B, Liu Y, et al. Infinite switching dynamic probabilistic network with Bayesian nonparametric learning. *IEEE Trans Signal Process*, 2022, 70: 2224–2238
- 29 Chung J Y, Kastner K L, Dinh L, et al. A recurrent latent variable model for sequential data. In: Proceedings of International Conference on Neural Information Processing Systems, 2015. 2980–2988
- 30 Li Z, Zhao Y, Han J, et al. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021. 3220–3230
- 31 Deng A, Hooi B. Graph neural network-based anomaly detection in multivariate time series. In: Proceedings of AAAI Conference on Artificial Intelligence, 2021, 4027–4035
- 32 Zhao H, Wang Y, Duan J, et al. Multivariate timeseries anomaly detection via graph attention network. In: Proceedings of Industrial Conference on Data Mining, 2020. 841–850
- 33 Chen W, Tian L, Chen B, et al. Deep variational graph convolutional recurrent network for multivariate time series anomaly detection. In: Proceedings of International Conference on Machine Learning, 2022. 3621–3633
- 34 Zhou M, Hannah L, Dunson D. Lawrence C: beta-negative binomial process and poisson factor analysis. In: Proceedings of International Conference on Artificial Intelligence and Statistics, 2012. 1462–1471
- 35 Graves A. Generating sequences with recurrent neural networks. 2013. ArXiv:1308.0850
- 36 Chung J, Ahn S, Bengio Y. Hierarchical multiscale recurrent neural networks. In: Proceedings of International Conference on Learning Representations, 2017
- 37 Guo D, Chen B, Lu R, et al. Recurrent hierarchical topic-guided RNN for language generation. In: Proceedings of International Conference on Machine Learning, 2020. 3810–3821
- 38 Chen W, Chen B, Liu Y, et al. Bidirectional recurrent gamma belief network for HRRP target recognition. *Signal Process*, 2021, 188: 108213
- 39 Zhou M Y, Cong Y L, Chen B. Augmentable gamma belief networks. *J Mach Learn Res*, 2016, 17: 5656–5699
- 40 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*, 1997, 9: 1735–1780
- 41 Kingma D P, Welling M. Auto-encoding variational Bayes. In: Proceedings of International Conference on Learning Representations, 2014
- 42 Rezende D J, Mohamed S, Wierstra D. Stochastic back-propagation and variational inference in deep latent Gaussian models. In: Proceedings of International Conference on Machine Learning, 2014
- 43 Guo D, Chen B, Chen W, et al. Variational temporal deep generative model for radar HRRP target recognition. *IEEE Trans Signal Process*, 2020, 68: 5795–5809
- 44 Zhou M Y, Cong Y L, Chen B. The poisson gamma belief network. In: Proceedings of International Conference on Neural Information Processing Systems, 2015. 3043–3051
- 45 Child R. Very deep VAEs generalize autoregressive models and can outperform them on images. In: Proceedings of International Conference on Learning Representations, 2021
- 46 Razavi A A, Oord A V D, Poole B, et al. Preventing posterior collapse with delta-VAEs. In: Proceedings of International Conference on Learning Representations, 2019
- 47 Siffer A, Fouque P A, Termier A, et al. Anomaly detection in streams with extreme value theory. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery, 2017. 1067–1075

# Stacked adversarial variational recurrent neural network for anomaly detection of multivariate time series

Wenchao CHEN<sup>1</sup>, Bowei FANG<sup>1</sup>, Liang DAI<sup>2</sup>, Bo CHEN<sup>1\*</sup>, Chang LIU<sup>2</sup> & Xiaonan ZHAO<sup>3\*</sup>

1. *National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China;*

2. *Institute of Information Engineering, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing 100093, China;*

3. *Shandong New Generation Information Industry Technology Research Institute, Ltd., Jinan 250100, China*

\* Corresponding author. E-mail: bchen@mail.xidian.edu.cn, zxnan891100214@163.com

**Abstract** To characterize the complex temporal dependence and stochasticity of multivariate time series (MTS) and achieve their unsupervised anomaly detection, which is crucial for managing service quality in industrial devices and Internet infrastructures. Our proposed method, called stacked adversarial variational recurrent neural network (SaVRNN), is a hierarchical probabilistic dynamical model that unifies various elements, including hierarchical generative models, stacked recurrent structure, and multilayer adversarial optimization within a Bayesian framework. The fundamental concept behind SaVRNN is to capture the typical patterns in input time series data by considering multilevel and long-range temporal dependencies through stacked recurrent structures and multilayer shape characteristics via hierarchical probabilistic generative operations. Subsequently, we utilize reconstruction probabilities to identify anomalies. To facilitate efficient inference, we introduce a unique upward-downward adversarial inference scheme, which accurately approximates latent variables' posterior distribution. However, given the challenges associated with balancing the training of the generator and discriminator in a multilayer adversarial network, SaVRNN can be challenging to train effectively. To address this issue, we have developed an optimization method based on conditional transport. Leveraging the hierarchical probabilistic structure inherent to SaVRNN, we propose a modified anomaly detection metric that combines the likelihood information from multiple layers. We evaluate the performance of SaVRNN through experiments conducted on two publicly available datasets and one real-world dataset. The results demonstrate the efficacy of SaVRNN for online anomaly detection.

**Keywords** anomaly detection, multivariate time series, adversarial variational Bayesian, variational RNN, conditional transport