控制与决策
Control and Decision

一种采用串行自编码器的时序数据异常检测方法

徐天慧, 郭强, 张彩明

引用本文:
徐天慧,郭强,张彩明. 一种采用串行自编码器的时序数据异常检测方法[J]. 控制与决策, 2023, 38(12): 3507−3515.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.0318

一种采用串行自编码器的时序数据异常检测方法

徐天慧[1], 郭 强[1,2,3†], 张彩明[2,3,4]

(1. 山东财经大学 计算机科学与技术学院，济南 250014；
2. 山东财经大学 山东省数字媒体技术重点实验室，济南 250014；
3. 山东省未来智能金融工程实验室，山东 烟台 264005；4. 山东大学 软件学院，济南 250101)

摘 要: 基于深度学习的时序数据异常检测模型大多采用循环神经网络或长短期记忆网络捕捉时序依赖性,并利用自编码器重构数据,进而实现时序数据的异常检测. 虽然此类检测模型实现了较高的异常检测率,但它们的网络结构复杂,导致模型的计算效率较低. 为提高模型的计算效率,提出一种基于串行自编码器的异常检测模型SAE-AD. 该模型仅包含两个结构简单的自编码器(AE_1和AE_2),其所含参数较少,且训练目标较为简单,从而加快了模型的计算效率.通过将自编码器AE_1和AE_2串行拼接,即AE_1的输出作为AE_2的输入,可有效提高AE_2的解码器对正常数据特征的解码能力,有助于提升模型的检测准确率. 实验结果表明,相较于其他新近提出的异常检测模型,SAE-AD模型具有更高的精确率、召回率和F_1值.

关键词: 深度学习；时序数据；异常检测；自编码器；数据重构；编码器；解码器

中图分类号: TP391 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.0318

引用格式: 徐天慧,郭强,张彩明. 一种采用串行自编码器的时序数据异常检测方法[J]. 控制与决策, 2023, 38(12): 3507-3515.

A serial autoencoders based method for detecting time series anomalies

XU Tian-hui[1], *GUO Qiang*[1,2,3†], *ZHANG Cai-ming*[2,3,4]

(1. School of Computer Science and Technology，Shandong University of Finance and Economics，Jinan 250014，China；
2. Shandong Key Laboratory of Digital Media Technology，Shandong University of Finance and Economics，Jinan 250014，China；3. Shandong Provincial Laboratory of Future Intelligence and Financial Engineering，Yantai 264005，China；
4. Software College，Shandong University，Jinan 250101，China)

Abstract: Aiming to detect time series anomalies, deep learning methods generally use the recurrent neural network or long short term memory to capture temporal dependency, and adopt autoencoder to reconstruct data. Although they work well for detecting anomalies, the network structures of these methods are complex, resulting in slow computational efficiency. In order to improve the computational efficiency, this paper proposes a method called serial autoencoders based anomaly detection (SAE-AD) which contains two autoencoders (AE_1 and AE_2) with simple structure. Due to the simplicity, there are a few training parameters and its training objectiv function is relatively simple, which speeds up the computation. In addition, the output of AE_1 is fed into AE_2 to improve the decoding ability of the decoder of AE_2. This way of serial training makes SAE-AD achieve better detection accuracy. Experiment results show that the proposed method has better precision, recall, F_1 score than several state-of-art anomaly detection methods.

Keywords: deep learning；time series；anomaly detection；autoencoder；data reconstruction；encoder；decoder

0 引 言

时序数据由随时间展开的序列值组成,这些序列值真实地记录着系统各个时刻的重要信息[1]. 时序数据分析旨在从大量的时序数据中提取潜在的有用信息,其主要任务包含聚类[2]、分类[3]和异常检测[4]等. 其中,异常检测的目标是检测出不服从正常数据模式的异常点,在网络入侵检测[5]、工业故障诊断[6]、人体行为异常检测[7]等领域均有着广泛的应用.

目前,研究人员已对时序数据的异常检测进行了深入研究[8-9]. 一种常用策略是通过度量数据之间的

收稿日期: 2022-03-01；**录用日期:** 2022-07-06.
基金项目: 国家自然科学基金项目(61873145, 61802229)；山东省自然科学省属高校优秀青年人才联合基金项目(ZR2017JL029)；山东省高等学校青创科技支持计划项目(2019KJN045).
责任编委: 胡清华.
†通讯作者. E-mail: guoqiang@sdufe.edu.cn.

邻近度实现异常检测. 由于异常数据不服从正常的数据模式,其特征往往与正常数据的特征存在明显差异,导致两者之间的邻近度较低. 在检测阶段,将与正常数据邻近度较低的待检数据判定为异常;否则,判定为正常数据. 这种基于邻近度度量的异常检测方法的关键在于选择一个合适的邻近度度量标准. 常见的度量标准有相似性[10]、距离[11]和密度[12]等.

尽管基于邻近度度量的方法有较低的误报率,但其性能对邻近度度量标准较为敏感,这限制了此类方法的应用. 鉴于深度神经网络强大的特征提取和表示能力,基于深度学习的异常检测方法受到了广泛关注[13]. 通过有效提取输入数据的特征以获得低维的特征表示,并利用该特征表示进行数据重构,再将重构误差与预先设定的阈值进行比较即可检测异常[14]. 深度神经网络能有效地提取出正常数据的特征,从而使正常数据取得较小的重构误差,异常数据的重构误差则较大[15]. 因此,若待检数据的重构误差高于阈值,则判定待检数据为异常;否则,判定为正常. 上述基于深度学习的异常检测方法通常借助自编码器(autoencoder, AE)[16]、循环神经网络(recurrent neural network, RNN)[17]及其变体提取数据特征. RNN及其变体的引入使检测方法能够捕捉数据的时序依赖性,在一定程度上提高了检测精度. 然而,导致检测方法的网络结构过于复杂,其计算效率较低[18].

为提高检测模型的计算效率和检测性能,本文提出一种结构简单的基于串行自编码器的异常检测模型(serial autoencoders based anomaly detection, SAE-AD). 该模型使用两个AE(AE_1和AE_2)重构输入数据,每个AE包含一个编码器(E_1/E_2)和一个解码器(D_1/D_2). 简单的网络结构以及模型中较少的参数缩短了其训练时间. SAE-AD通过串行训练AE_1和AE_2,将AE_1的输出结果输入至AE_2,以提高D_2对正常数据特征的解码能力,并利用E_1和D_2计算待检数据的重构误差. 该结构有利于放大异常数据的重构误差,从而提升模型的异常检测率. 本文在公共数据集上将SAE-AD模型与5种新近提出的检测模型分别进行比较,以验证本文模型的有效性.

1 相关工作

得益于深度学习的快速发展,研究人员提出了许多基于深度学习的异常检测模型[13],这些模型的性能依赖于所用网络提取数据特征的能力. 一种常用的特征提取网络结构是AE,它由一个编码器与一个解码器组成. 一方面,编码器在潜在空间中提取输入数据的特征并给出该特征的低维表示;另一方面,解码器对这些低维表示进行升维以重构输入数据[19]. 在编码过程中,数据中的部分干扰信息会受到抑制,因而编码器能够较好地获取正常数据的特征表示. 在解码器的作用下,AE可重构出还原度较高的正常数据,而对异常数据的重构误差则较大. Zhou等[16]将AE的重构误差记为异常得分,根据此得分的大小实现了时序数据的异常检测.

Zong等[20]借助AE重构数据的能力,提出了用于异常检测的深度自编码高斯混合模型[1]. 该模型使用AE获取数据的低维特征表示和重构表示,并利用高斯混合模型对特征表示和重构误差进行概率估计,通过选择合适的阈值,将概率估计高于阈值的数据判定为异常. 尽管该模型的检测精度较AE有所提升,但却忽略了时序数据之间固有的时序依赖性.

为捕捉时序依赖性,Cui等[17]利用RNN在隐藏层之间建立全连接,用于保存和传递时序状态信息. 然而,随着时间步的增加,其更新参数的过程会出现梯度消失或梯度爆炸的问题[21]. 针对该问题,将长短期记忆网络(long short term memory, LSTM)[22]引入细胞状态以保存信息,并利用门结构有选择地抑制或增强细胞状态中的信息. 上述改动使得LSTM可有效捕捉时序的长期依赖关系[23]. 基于此,文献[24]在编码器和解码器中引入LSTM网络对时序的长期依赖进行建模,并根据数据重构误差的大小检测异常.

除了上述AE以外,变分自编码器(variational autoencoder, VAE)也常被用于时序数据的异常检测[25-26]. VAE利用编码器建立原始输入数据的变分推断,生成隐变量的概率分布,并使用此分布对隐变量进行采样,然后再对采样后的隐变量进行解码得到输入数据的估计[27]. 基于LSTM的VAE网络模型[28]使用LSTM替代VAE中的前馈神经网络来建模输入数据的时序依赖性,并利用VAE生成输入数据的概率估计. 它将异常得分定义为待检数据相对于概率估计的负对数似然,从而检测时序异常. 虽然该模型可建模输入数据的时序依赖,但却忽视了隐变量之间的依赖性. 为此,Su等[5]引入随机变量连接技术以捕捉隐变量之间的依赖性,并计算待检数据服从VAE所生成的数据分布的概率,进而提高了检测精度.

上述利用VAE进行异常检测的模型实现了较高的异常检测率,然而,这些模型所含参数较多,大量参数的迭代更新导致其训练速度较低. 为提高检测模型的训练速度,Audibert等[18]提出了一种所含参数较少的无监督异常检测模型(unsupervised

anomaly detection, USAD), 它由两个共享编码器的AE构成. 其检测异常的关键在于通过对抗训练这两个AE放大异常数据的重构误差, 表现出较好的检测性能. 受此启发, 本文提出无监督的SAE-AD模型. 不同于UASD采用共享编码器的策略, 本文模型串行拼接两个AE, 可进一步放大异常数据的重构误差.

2 本文模型

2.1 问题描述

给定一组长度为n的多元时间序列$X=\{X_1, X_2, \ldots, X_n\}$, 其$t(1 \leqslant t \leqslant n)$时刻的观测值是$m$维的向量$X_t = [x_1, x_2, \ldots, x_m]$, 它记录了此时的系统状态信息. 时序数据异常检测的目标是判断时间序列X中是否存在异常, 若存在, 则定位该异常. 为提高检测模型的鲁棒性, 本文使用长度为k的滑动时间窗口对时序数据X进行分割[29], 进而得到子序列集合$W = \{W_1, W_2, \ldots, W_n\}$, 其中子序列$W_t = \{X_{t-k+1}, X_{t-k+2}, \ldots, X_t\}$为$t$时刻的窗口数据. 对数据$X_t$的异常检测可以转化为对子序列$W_t$的异常检测, 即通过构建异常检测器$f(W_t)$判断待检序列$W_t$是否含异常. 若含异常, 则检测器输出1, 否则输出0. 将检测器$f(W_t)$定义为

$$f(W_t) = \begin{cases} 0, \ \mathrm{Score}(W_t) < \lambda; \\ 1, \ \mathrm{Score}(W_t) \geqslant \lambda. \end{cases} \tag{1}$$

其中: $\mathrm{Score}(W_t)$记为时刻t的异常得分值; λ为预先设置的阈值, 用于判定子序列W_t是否为异常序列.

2.2 自编码器

AE是一种应用广泛的无监督神经网络, 其主要功能是提取数据的特征并重构数据. 图1给出了AE的基本结构, 其主要由结构对称的编码器与解码器两个部分组成[19].

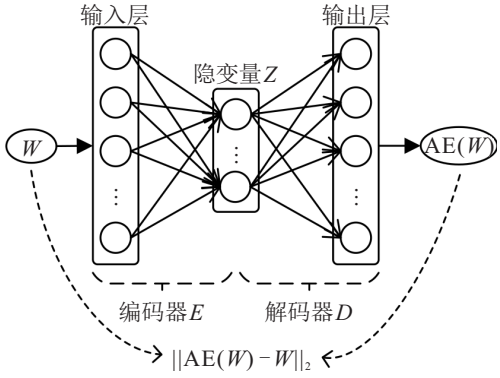


图1　自编码器

编码器能够提取输入层数据W的特征, 并给出特征的低维表示, 即隐变量Z. 解码器对Z进行解码, 进而得到数据的重构表示$\mathrm{AE}(W)$. 上述编解码过程可以描述为

$$Z = E(W) = \sigma(a_E W + b_E), \tag{2}$$

$$\mathrm{AE}(W) = D(Z) = \sigma(a_D Z + b_D). \tag{3}$$

其中: a_E、b_E和a_D、b_D分别为编码器和解码器的权重、偏置, σ为非线性激活函数.

为获得有效的数据特征表示, 通常将AE的训练目标设置为最小化输入数据的重构误差, 即最小化$\mathrm{AE}(W)$与W之间的差异

$$\min_{\mathrm{AE}} \|\mathrm{AE}(W) - W\|_2, \tag{4}$$

其中$\|\cdot\|_2$为L_2范数. 在编码过程中, 通过对输入数据降维以抑制数据中的干扰信息, 使得隐变量能够较好地描述正常数据的特征, 从而保证该特征解码输出的误差较小. 理想情况下, 重构输出$\mathrm{AE}(W)$与输入数据W完全一致. 然而, 对于异常数据, 经由正常数据训练后的AE在编码过程中, 会在一定程度上抑制异常信息, 从而无法较好地捕捉异常数据的特征, 这导致异常数据的重构输出与原始输入存在较大差异.

为进一步提高AE的编码能力, 深度AE通过加深AE的网络层数使编码器可抑制更多的干扰信息, 以得到更有效的数据特征表示[10]. 图2给出了深度AE的一种网络结构, 该深度AE的编码器和解码器均由3个全连接层构成. 通过极小化重构输出与输入数据之间的差异实现较为准确的正常数据重构.

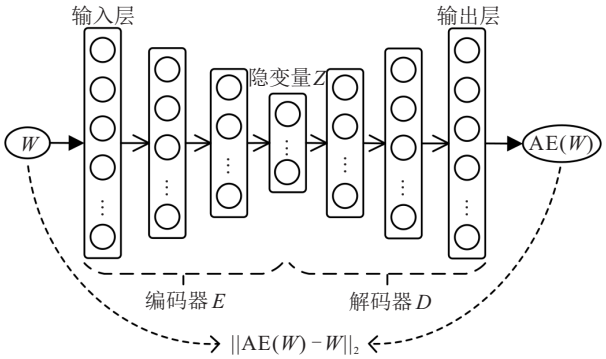


图2　深度自编码器

2.3 串行自编码器

虽然使用AE可检测时序异常, 但是其取得的检测精度较低[18]. 针对该问题, 本文提出SAE-AD检测模型, 其核心在于放大异常数据的重构误差. 该模型由两个结构相同的AE(AE_1和AE_2)构成, 每个AE各包含一个编码器(E_1/E_2)和一个解码器(D_1/D_2). 编码器和解码器均由3个全连接层构成. 将输入层和输出层的大小设置为$k \cdot m$. 其中: k为时间窗口的大小, m为输入数据的维度. 将隐变量Z的维度定义为h, 编码器中全连接层的大小分别为前一层的$1/2$, 解

码器与编码器结构对称,易知模型需要优化的参数总量为$k\cdot m(5/2k\cdot m+h+5)+2h$,较少的参数有助于模型实现较为高效的异常检测.

本文将AE_1的解码输出作为AE_2的编码输入,用于提高D_2对正常数据特征的解码能力. 在异常检测阶段,将E_1与D_2拼接为异常得分计算器AE_3,并使用异常检测器$f(W_t)$判定是否为异常.

2.3.1 学习阶段

图3(a)给出了SAE-AD模型的学习过程和检测过程.

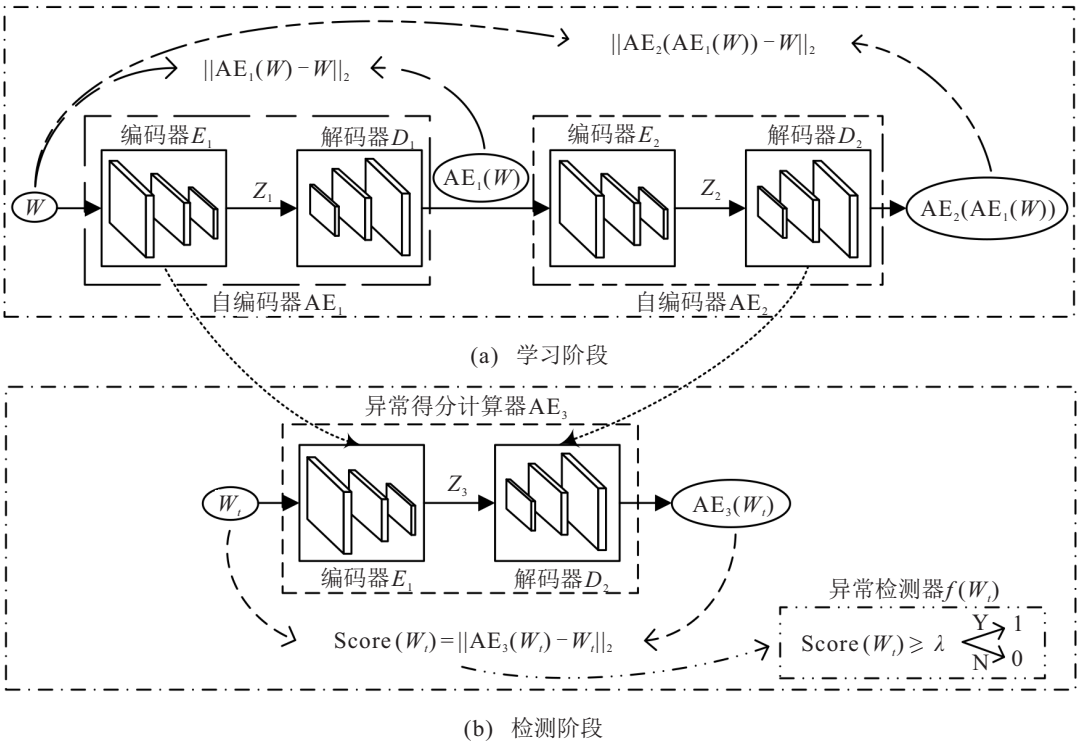


图3 SAE-AD模型的学习过程和检测过程

为了能让AE_1和AE_2均提取出有用的数据特征,得到有效的低维特征表示,本文分别将它们的训练目标函数定义为

$$\min_{\mathrm{AE}_1}\|\mathrm{AE}_1(W)-W\|_2, \tag{5}$$

$$\min_{\mathrm{AE}_2}\|\mathrm{AE}_2(\mathrm{AE}_1(W))-W\|_2. \tag{6}$$

使用式(5)训练AE_1以获得输入序列W的低维特征表示Z_1. 在E_1的降维作用下,Z_1中包含正常数据的共性信息,同时抑制了部分干扰信息. 因此,D_1对Z_1进行解码得到的重构输出$\mathrm{AE}_1(W)$与正常序列W的重构误差较小. 对于AE_2,本文将AE_1的重构输出$\mathrm{AE}_1(W)$作为其编码输入. 这使得E_2进一步抑制干扰信息,而且Z_2比Z_1能够更准确地描述正常数据特征. 因此,利用D_2对Z_2进行解码重构出的$\mathrm{AE}_2(\mathrm{AE}_1(W))$所含干扰信息较$\mathrm{AE}_1(W)$更少,同时增强了$D_2$对正常数据特征的解码能力. 此外,为防止$\mathrm{AE}_2$在获取低维特征表示过程中丢失过多信息,将$\mathrm{AE}_2$的训练目标函数设置为极小化$\mathrm{AE}_2(\mathrm{AE}_1(W))$与输入$W$之间的差异,而非$\mathrm{AE}_2(\mathrm{AE}_1(W))$与$\mathrm{AE}_1(W)$之间的差异.

2.3.2 检测阶段

在AE_1和AE_2的编解码作用下,SAE-AD模型会得到待检序列W_t的重构表示$\mathrm{AE}_2(\mathrm{AE}_1(W_t))$. 由于AE的编解码过程可抑制部分干扰信息,当$W_t$含异常数据时,$\mathrm{AE}_1$和$\mathrm{AE}_2$会去除部分异常信息,进而导致序列$W_t$的重构误差较大. 然而,当$W_t$为正常序列时,由于数据中往往存在噪声干扰,而且有些噪声会被AE_1和AE_2去除,W_t与$\mathrm{AE}_2(\mathrm{AE}_1(W_t))$之间也会存在一定的重构误差,这会影响检测器对异常的判定.

为放大异常序列的重构误差,同时缩小正常序列的重构误差,SAE-AD模型将AE_1的编码器E_1和AE_2的解码器D_2构成异常得分计算器(记为AE_3),如图3(b)所示. 其仅含一个编码器E_1能够防止抑制过多的干扰信息. 因此,相对于使用串行AE_1和AE_2作为异常得分计算器,AE_3缩小了正常数据的重构误差. 与此同时,对于异常序列,本文使用对正常数据特征解码能力较强的D_2进行解码,能够在一定程度上放大异常数据的重构误差. 这是因为在模型的学习

过程中, Z_2 比 Z_1 包含的干扰信息特征更少, 所以 D_2 比 D_1 能够更好地对正常数据特征进行重构表示.

将待检序列 W_t 的异常得分记为

$$\mathrm{Score}(W_t) = \|\mathrm{AE}_3(W_t) - W_t\|_2, \tag{7}$$

并使用异常检测器 $f(W_t)$(见式(1))比较异常得分 $\mathrm{Score}(W_t)$ 与预先设置的阈值 λ 的大小, 进而确定待检序列是否异常.

2.3.3 算法过程

本文所提SAE-AD模型的具体实现步骤总结如下:

1) 模型学习阶段.

step 1: 数据预处理. 使用滑动窗口技术将时序数据 $X=\{X_1, X_2, \ldots, X_n\}$ 分割为 $W = \{W_1, W_2, \ldots, W_n\}$.

step 2: 串行训练. W 作为 AE_1 的输入, 将其输出结果 $\mathrm{AE}_1(W)$ 输入至 AE_2, 并获得重构输出 $\mathrm{AE}_2(\mathrm{AE}_1(W))$, 以实现 AE_1 和 AE_2 的串行训练.

step 3: 参数学习. 使用式(5)和(6)优化 AE_1 和 AE_2 中的参数 a_E、b_E 和 a_D、b_D.

2) 异常检测阶段.

step 1: 待检数据预处理. 采用模型训练阶段相同的数据预处理策略将待检数据 X 分割为 $W_t (t = 1, 2, \ldots, n)$.

step 2: 数据重构. 利用 AE_3 对待检序列 W_t 进行重构, 进而输出重构结果 $\mathrm{AE}_3(W_t)$.

step 3: 计算异常得分. 由式(7)计算 W_t 的重构误差, 并得到该序列的异常得分 $\mathrm{Score}(W_t)$.

step 4: 异常检测. 将 $\mathrm{Score}(W_t)$ 与阈值 λ 加以比较, 若得分超出阈值, 则将 X_t 判定为异常, 反之判定为正常数据.

3 实验与结果分析

3.1 数据集

本文在5个公开数据集SWaT、WADI、SMD、SMAP和MSL上对检测模型的性能进行评估. SWaT和WADI数据集为单实体数据集, 其他3个数据集为多实体数据集. SWaT数据集记录污水净化厂中与水处理过程相关的物理属性以及测试台上的网络流量数据, 包含7天正常操作的数据和4天异常操作时的数据[30]. WADI数据集是SWaT数据集的拓展, 其包含14天的正常数据和2天的异常数据[30]. SMD数据集是由一家互联网公司收集并公开的, 其数据记录了28台服务器在5周内运行的数据, 数据维度为38维[31]. 每台服务器的数据均被分为2个大小相等的子集, 分别作为训练集和测试集. SMAP和MSL数据集是来自NASA专家标记的真实遥感数据集[31]. SMAP数据集含有55个实体, 每个实体有25维度的数据; MSL数据集包含27个实体, 每个实体的数据维度为55. 表1列出了上述5个数据集用于训练和测试的数据量、数据维度信息以及异常比例.

表1　数据集的数据量、维度和异常比例信息

数据集	训练数据	测试数据	数据维度	异常比例/%
SWaT	496 800	449 919	51	11.98
WADI	1 048 571	172 801	123	5.99
SMD	708 405	708 420	28×38	4.16
SMAP	135 183	427 617	55×25	13.13
MSL	58 317	73 729	27×55	10.72

3.2 评估标准

为评价各种异常检测模型的性能, 对于单实体数据集, 本文采用精确率(precision, P)、召回率(recall, R)和综合评价指标作为评估标准, 分别定义如下:

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \ F_1 = 2 \cdot \frac{P \cdot R}{P + R}.$$

其中: TP为正类实例中被预测为正类的数量, FP和TN分别为负类被预测成正类和负类的数量. 精确率反映预测为异常的数据中真实异常所占比例, 体现检测模型对异常检测的准确度; 召回率表示异常数据被检测到的比例, 反应模型对异常检测的全面性; F_1 值是结合精确率和召回率两个方面的综合性能. 为保证评价的公平性, 在每个数据集上对各模型分别进行10次实验, 并取其指标均值.

对于多实体数据集, 除上述3种评估标准外, 采用了 F_1^* 值以评估模型的检测性能. F_1^* 由数据集中所有实体的 P、R 均值计算得到, 其定义[5]为

$$F_1^* = 2 \cdot \frac{\bar{P} \cdot \bar{R}}{\bar{P} + \bar{R}},$$

其中 $\bar{P}$ 和 $\bar{R}$ 分别为数据集中各实体的 P 和 R 均值.

3.3 实验设置与对比方法

将本文模型SAE-AD与5种新近提出的基于深度学习的检测模型进行对比, 包括自编码器(AE)、深度自编码高斯混合模型(DAGMM)[20]、随机循环神经网络模型(OmniAnomaly)[5]、基于短期记忆网络的变分自编码器(LSTM-VAE)[28]和无监督异常检测模型(USAD)[18]. DAGMM、OmniAnomaly、LSTM-VAE和USAD模型采用公开源代码及默认参数设置, AE和本文模型则由Pytorch框架实现.

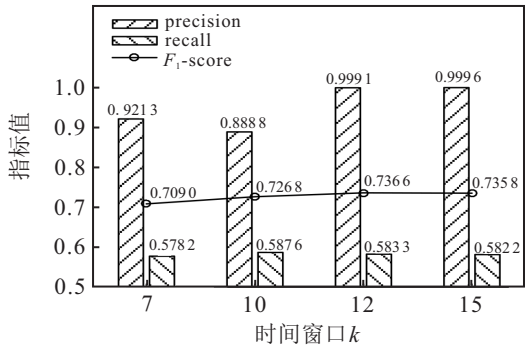
采用Adam优化器训练SAE-AD模型. 在模型训练和测试过程中, 均使用GTX 1080 GPU进行加速. 在SWaT、WADI、SMAP和MSL数据集上, 本文模型的迭代训练次数epochs为125. 对于SMD数据集,

本文将epochs设为250. 在异常检测阶段,对可能的阈值λ取值进行测试,经验地选取当模型综合性能F_1最优时的λ大小,并记录此时的指标值. 在SWaT、WADI、SMD、SMAP和MSL数据集上,λ取值分别为0.843 8、0.878 4、0.099 4、0.087 8和0.255 8.

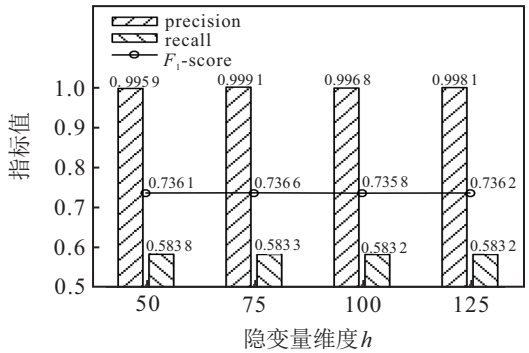
3.4 结果分析

3.4.1 参数设置

SAE-AD模型包含两个重要参数,分别是时间窗口大小k和隐变量Z_3的维度h. 为分析k对本文模型检测性能的影响,首先固定h为75,再分别将窗口大小设置为k = 7、10、12和15,并在SWaT数据集上进行异常检测. 图4(a)给出了选取不同窗口大小时SAE-AD模型的检测结果.



(a) 滑动窗口长度



(b) 隐变量维度

图4 不同参数设置对实验结果的影响

从图4(a)中可以看出,尽管k = 12时该模型取得了次高的精准率和召回率,但是有最高的F_1值为0.736 6,比最低值明显高出2.74 %. 由此说明k为12时,本文模型存在一定的漏检和误检,但是其综合性能较高. 因此,在SWaT数据集上,选取长度为12的时间窗口对数据进行分割. 此外,对于WADI和SMD数据集,经验地将k分别设置为10和5;对于SMAP和MSL数据集,将其设置为12.

为分析隐变量Z_3的维度h对模型检测性能的影响,将k固定为12,并选取不同的h = 50、75、100和125进行实验. 图4(b)给出了在SWaT数据集上的检测结果. 观察到,当h = 75时SAE-AD模型取得的召回率次高,比最高值低0.05 %,这说明模型存在漏检情况. 尽管如此,此时该模型取得了最高的精准率和F_1值,分别比最低值高出0.32 %和0.08 %,充分说明h为75时,本文模型检测异常的准确度和综合性能较高. 在SWaT、WADI、SMAP和MSL数据集上,经验地将h均设置为75. 由于SMD数据集对该参数较为敏感,将h调整为38以实现较高的检测性能. 在上述5个数据集上,SAE-AD模型所用时间窗口大小k、隐变量Z_3的维度h、模型迭代训练次数epochs和阈值λ取值如表2所示.

表2 SAE-AD模型在各数据集上的参数设置

数据集	k	h	epochs	λ
SWaT	12	75	125	0.843 8
WADI	10	75	125	0.878 4
SMD	5	38	250	0.099 4
SMAP	12	75	125	0.087 8
MSL	12	75	125	0.255 8

3.4.2 对比分析

表3列出了不同异常检测模型在单实体数据集SWaT和WADI上的检测结果,最高和次高值由粗体标出.

表3 检测模型在SWaT和WADI数据集上的实验结果

检测模型	SWaT			WADI		
	P	R	F_1	P	R	F_1
AE	0.238 4	**0.724**	0.358 5	**0.997 4**	0.1497	**0.260 3**
DAGMM	0.679 7	**0.629 9**	0.653 8	0.368 2	0.122 1	0.183 4
LSTM-VAE	**0.993 1**	0.589 9	**0.740 2**	**0.997 4**	0.149 7	**0.260 3**
OmniAnomaly	0.974 1	0.418 7	0.585 6	0.09 46	**0.966 3**	0.172 3
USAD	0.966 6	0.585 7	0.729 4	**0.996 1**	0.149 7	**0.260 4**
SAE-AD	**0.999 1**	0.583 3	**0.736 6**	**0.997 4**	**0.149 8**	**0.260 4**

从表3中可以看出,在SWaT数据集上,虽然SAE-AD模型的召回率相对较低,但其取得了最高的精确率0.999 1和次高的F_1值0.736 6,分别较AE高出0.760 7和0.378 1. 这说明尽管本文模型存在一定程度的漏检,但其对异常的检测更为准确,在综合性能方面也具有一定的优势,且明显优于AE模型. 这是

因为相较于仅使用一个AE重构数据,SAE-AD通过串行两个AE能抑制更多的干扰信息,从而提升模型的鲁棒性. 对于WADI数据集,DAGMM模型的各项指标均较低. 这是由于该模型的输入为单一数据,而非时序子序列,导致其无法捕捉时序依赖性.

本文模型采用滑动窗口技术将子序列作为模型输入,能够建立时序关系. 因此,本文模型取得的精准率和F_1值更高,实现了较好的检测准确度和综合性能. 此外,还观察到在该数据集上OmniAnomaly模型的召回率最高为0.966 3,而其精准率却明显低于其他模型. 产生这种现象的原因在于该模型对检测异常时所用阈值较为敏感.

不同检测模型在多实体数据集SMD、SMAP和MSL上的检测结果如表4所示.

表 4　检测模型在SMD、SMAP和MSL数据集上的实验结果

检测模型	SMD				SMAP				MSL			
	P	R	F_1	F_1^*	P	R	F_1	F_1^*	P	R	F_1	F_1^*
AE	0.590 6	0.486 7	0.456 4	0.533 7	**0.389 6**	0.637 7	**0.364 9**	0.483 7	0.306 6	0.637 9	**0.340 9**	0.414 1
DAGMM	0.114 5	0.466 9	0.130 3	0.183 9	0.124 9	0.423 1	0.137 1	0.192 8	0.158 9	0.331 5	0.164 3	0.214 8
LSTM-VAE	**0.595 6**	0.498 4	0.460 5	0.542 7	0.380 2	0.624 9	0.351 6	0.472 7	**0.316 5**	0.615 0	0.334 2	**0.417 9**
OmniAnomaly	**0.732 9**	**0.942 9**	**0.749 7**	**0.824 7**	0.116 4	**0.999 9**	0.196 2	0.208 5	0.125 3	**0.999 9**	0.195 8	0.222 7
USAD	0.579 8	0.522 8	0.465 2	0.549 9	0.389 3	0.640 4	0.364 6	**0.484 3**	0.309 4	**0.638 8**	0.337 8	0.416 8
SAE-AD	0.568 9	**0.544 3**	**0.473 5**	**0.556 4**	**0.405 6**	**0.686 5**	**0.382 4**	**0.509 9**	**0.331 5**	0.632 6	**0.357 0**	**0.435 0**

由表4可看出,对于SMD数据集,OmniAnomaly模型的各项指标均最高,这说明该模型所用随机变量连接技术能捕捉隐变量之间的依赖性,有助于提升方法的检测性能. 虽然SAE-AD模型的精确率较低为0.568 9,但其取得了次高的召回率、F_1值和F_1^*值. 这说明在该数据集上,SAE-AD模型存在一定的误检现象,但其对异常的查全率和整体性能较高. 无论SMAP还是MSL数据集,SAE-AD模型检测异常所取得的精确率、F_1值和F_1^*值明显高于DAGMM和OmniAnomaly模型,如在MSL数据集上,SAE-AD的F_1^*值比DAGMM和OmniAnomaly分别高出0.220 2和0.212 3. 由此可知,在这两个数据集上,本文模型通过串行训练两个AE提高了D_2对正常数据特征的解码能力,再利用E_1和D_2计算异常得分能够放大异常数据的重构误差. 因此,在异常检测的准确度和综合性能方面具有较好的表现. 此外,还观察到在这3个多实体数据集上,本文模型的大部分指标高于USAD模型. 据此说明,相较于UASD模型对抗训练两个AE的方式,SAE-AD模型通过串行训练实现了更好的检测性能.

由表3和表4中的实验结果可以看出: AE和DAGMM模型的检测性能较差; LSTM-VAE、USAD和OmniAnomaly模型的精确率和召回率比AE和DAGMM模型有所提升; SAE-AD模型的综合性能最优. 为比较模型的计算效率,本文对比了不同模型的时间复杂度. 由于LSTM-VAE和OmniAnomaly模型的实现分别依赖于LSTM和门控循环单元(GRU),且LSTM和GRU的优化需消耗较长计算时间[18],LSTM-VAE和OmniAnomaly的时间复杂度较高. USAD和SAE-AD模型仅含有AE结构,因此它们的计算复杂度低于LSTM-VAE和OmniAnomaly模型. USAD和SAE-AD模型的结构区别在于USAD比SAE-AD少一个编码器. 尽管如此,USAD模型采用的对抗训练思想使得每次迭代训练时,其编码器中的参数需更新两次. 这意味着USAD所需更新参数量与SAE-AD模型相等,均为$k \cdot m(5/2k \cdot m + h + 5) + 2h$. 由于USAD模型的训练目标函数相对较为复杂,SAE-AD模型的串行训练目标函数更为简单,这在一定程度上缩短了本文模型的训练时间,有利于提高其计算效率.

综上所述,不论是单实体数据集,还是多实体数据集,本文模型不仅具有较高的精准率、召回率、F_1值和F_1^*值,而且模型的训练时间较短.

3.4.3　消融分析

为探究AE的数量s对模型的异常检测性能的影响,本文在SWaT、WADI、SMD、SMAP和MSL数据集上,选取不同的s值分别进行实验. 将3个AE串行训练组成SAE-AD3模型,并将第1个AE的编码器与第3个AE的解码器构成异常得分计算器,进而实现异常检测. 此外,为验证本文模型中异常得分计算器的结构的有效性,利用SAE-AD模型中的E_2和D_2重构待检序列计算异常得分,并对其(记为E2D2-AD模型)进行测试.

表5列出了AE($s = 1$)、SAE-AD($s = 2$)、SAE-AD3($s = 3$)和E2D2-AD模型在单实体数据集上的检测结果.

表5 检测模型在SWaT和WADI数据集上的实验结果

检测模型	SWaT			WADI		
	P	R	F_1	P	R	F_1
AE ($s=1$)	0.2384	**0.7224**	0.3585	0.9974	0.1497	0.2603
SAE-AD ($s=2$)	**0.9991**	0.5833	**0.7366**	**0.9974**	**0.1498**	**0.2604**
SAE-AD3 ($s=3$)	0.9989	0.5781	0.7324	0.9974	0.1497	0.2603
E2D2-AD	0.4488	0.7074	0.5492	0.9974	0.1497	0.2603

在SWaT数据集上,尽管AE取得了最大的召回率,但是SAE-AD模型的精准率和F_1值最高,明显高出AE模型0.760 7和0.378 1. 这说明本文模型的检测准确率和综合性能明显优于AE模型,并验证了利用串行训练的方式重构数据能够提升方法的检测性能. 此外,SAE-AD模型的F_1值较E2D2-AD模型高出0.187 4,充分体现出本文模型结构的有效性. 对于WADI数据集,SAE-AD取得了最高的F_1值为0.260 4,略高于其他模型,说明该模型具有相对较好的综合检测性能.

上述4种检测模型在多实体数据集上的检测结果如表6所示.

表6 检测模型在SMD、SMAP和MSL数据集上的实验结果

检测模型	SMD				SMAP				MSL			
	P	R	F_1	F_1^*	P	R	F_1	F_1^*	P	R	F_1	F_1^*
AE ($s=1$)	0.5906	0.4867	0.4564	0.5337	0.3896	0.6377	0.3649	0.4837	0.3066	**0.6379**	0.3409	0.4141
SAE-AD ($s=2$)	0.5689	**0.5443**	**0.4735**	**0.5564**	**0.4056**	**0.6865**	**0.3824**	**0.5099**	0.3315	0.6326	**0.3570**	**0.4350**
SAE-AD3 ($s=3$)	**0.6253**	0.4852	0.4090	0.5464	0.3965	0.6799	0.3741	0.5009	0.3315	0.6319	0.3567	0.4349
E2D2-AD	0.4255	0.4298	0.2848	0.4276	0.3966	0.6767	0.3822	0.5001	**0.3278**	0.6311	0.3529	0.4315

由表6不难看出,对于SMD数据集,SAE-AD3模型取得的精准率最高,然而SAE-AD模型的召回率、F_1值和F_1^*值最高,分别高出SAE-AD3模型0.059 1、0.064 5和0.01. 对于SMAP数据集,本文模型取得了最高指标值. 由此可知,该模型能够更有效地检测异常. 对于MSL数据集而言,尽管SAE-AD取得的精准率和召回率不是最高的,但是其F_1值和F_1^*值最高. 这说明相较于其他模型,该模型的综合性能最佳. 根据上述分析,$s=2$时串行自编码器的异常检测性能相对较高. 此外,在上述5个数据集上,SAE-AD模型的F_1值和F_1^*值均高于E2D2-AD模型. 尤其是在SMD数据集上,本文模型的F_1值和F_1^*值比E2D2-AD模型高出0.188 7和0.128 8,这足以验证本文模型结构的有效性.

4 结 论

本文借助AE提取特征和重构数据的能力,提出了一种有效的时序数据异常检测模型SAE-AD,其核心是通过串行训练两个AE (AE_1和AE_2)来提取正常数据共性的特征,并抑制数据中的干扰信息,从而提高了AE_2的解码器D_2对正常数据特征的解码能力. 在此基础上,利用AE_1的编码器E_1和D_2对待检数据进行重构,有助于放大异常数据的重构误差. 实验结果表明,本文模型取得了较高的精准率、召回率和F_1值,表现出较好的异常检测性能.

参考文献(References)

[1] Esling P, Agon C. Time-series data mining[J]. ACM Computing Surveys, 2012, 45(1): 1-34.

[2] Li H L, Liu Z C. Multivariate time series clustering based on complex network[J]. Pattern Recognition, 2021, 115: 107919.

[3] Li H L, Jia R Y, Wan X J. Time series classification based on complex network[J]. Expert Systems With Applications, 2022, 194: 116502.

[4] Blázquez-García A, Conde A, Mori U, et al. A review on outlier/anomaly detection in time series data[J]. ACM Computing Surveys, 2022, 54(3): 1-33.

[5] Su Y, Zhao Y J, Niu C H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]. KDD'19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, 2019: 2828-2837.

[6] 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.)

[7] 张晓平, 纪佳慧, 王力, 等. 基于视频的人体异常行为识别与检测方法综述[J]. 控制与决策, 2022, 37(1): 14-27.
(Zhang X P, Ji J H, Wang L, et al. Overview of video based human abnormal behavior recognition and detection

methods[J]. Control and Decision, 2022, 37(1): 14-27.)

[8] Chandola V, Banerjee A, Kumar V. Anomaly detection[J]. ACM Computing Surveys, 2009, 41(3): 1-58.

[9] 苏江军, 董一鸿, 颜铭江, 等. 面向复杂网络的异常检测研究进展[J]. 控制与决策, 2021, 36(6): 1293-1310.
(Su J J, Dong Y H, Yan M J, et al. Research progress of anomaly detection for complex networks[J]. Control and Decision, 2021, 36(6): 1293-1310.)

[10] Li H L. Time works well: Dynamic time warping based on time weighting for time series data mining[J]. Information Sciences, 2021, 547: 592-608.

[11] Li S, Xie Y, Dai H J, et al. Scan B-statistic for kernel change-point detection[J]. Sequential Analysis, 2019, 38(4): 503-544.

[12] Zhang L W, Lin J, Karim R. Adaptive kernel density-based anomaly detection for nonlinear systems[J]. Knowledge-Based Systems, 2018, 139: 50-63.

[13] Pang G S, Shen C H, Cao L B, et al. Deep learning for anomaly detection[J]. ACM Computing Surveys, 2022, 54(2): 1-38.

[14] Gong D, Liu L Q, Le V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019: 1705-1714.

[15] Rushe E, Namee B M. Anomaly detection in raw audio using deep autoregressive networks[C]. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing. Brighton, 2019: 3597-3601.

[16] Zhou C, Paffenroth R C. Anomaly detection with robust deep autoencoders[C]. KDD'17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, 2017: 665-674.

[17] Cui Q, Wu S, Liu Q, et al. MV-RNN: A multi-view recurrent neural network for sequential recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(2): 317-331.

[18] Audibert J, Michiardi P, Guyard F, et al. USAD: UnSupervised anomaly detection on multivariate time series[C]. KDD'20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, 2020: 3395-3404.

[19] 袁非牛, 章琳, 史劲亭, 等. 自编码神经网络理论及应用综述[J]. 计算机学报, 2019, 42(1): 203-230.
(Yuan F N, Zhang L, Shi J T, et al. Theories and applications of auto-encoder neural networks: A literature survey[J]. Chinese Journal of Computers, 2019, 42(1): 203-230.)

[20] Zong B, Song Q, Min M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]. Proceedings of the 6th International Conference on Learning Representations. Vancouver, 2018: 1-12.

[21] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.

[22] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[23] Ergen T, Kozat S S. Unsupervised anomaly detection with LSTM neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(8): 3127-3141.

[24] Malhotra P, Ramakrishnan A, Anand G, et al. LSTM-based encoder-decoder for multi-sensor anomaly detection[J/OL]. 2016, arXiv: 1607.00148.

[25] An J, Cho S. Variational autoencoder based anomaly detection using reconstruction probability[J]. Special Lecture on IE, 2015, 2(1): 1-18.

[26] Soelch M, Bayer J, Ludersdorfer M, et al. Variational inference for on-line anomaly detection in high-dimensional time series[J/OL]. 2016, arXiv: 1602.07109.

[27] Zhang H B, Wong R K, Chu V W. Hybrid variational autoencoder for recommender systems[J]. ACM Transactions on Knowledge Discovery from Data, 2022, 16(2): 1-37.

[28] 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.

[29] Aminikhanghahi S, Wang T H, Cook D J. Real-time change point detection with application to smart home time series data[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(5): 1010-1023.

[30] Mathur A P, Tippenhauer N O. SWaT: A water treatment testbed for research and training on ICS security[C]. 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater). Vienna, 2016: 31-36.

[31] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11(12): 3371-3408.

作者简介

徐天慧 (1998–), 女, 硕士生, 从事数据挖掘、异常检测等研究, E-mail: xth0606@126.com;

郭强 (1979–), 男, 教授, 博士, 从事计算机视觉、数据挖掘等研究, E-mail: guoqiang@sdufe.edu.cn;

张彩明 (1955–), 男, 教授, 博士, 从事计算机图形学、计算机视觉、医学影像处理和时序数据预测等研究, E-mail: czhang@sdu.edu.cn.