Transformer-Encoder-Based Anomaly Detection Algorithm in Multi-Sensor System*

LI Yan[1], NI Xiaoyu[2], CHEN Yu[3*]
(1.School of Mechanical and Electrical Engineering, Zhangjiakou Vocational and Technical College, Zhangjiakou Hebei 075000, China;
2.School of Mechanical Engineering, Hebei University of Architecture, Zhangjiakou Hebei 075000, China;
3.Academic Affairs Office, Zhangjiakou Open University, Zhangjiakou Hebei 075000, China)

Abstract: Multi-sensor anomaly detection plays a crucial role in several applications, including industrial monitoring, network-intrusion detection. However, the task poses significant challenges due to the presence of massive unlabeled data, and it is difficult to identify normal patterns in the spatio-temporal data. Therefore, Transformer-encoder-based anomaly detection (TEAD) is proposed. TEAD algorithm does not rely on labed data and belongs to unsupervised learning. It combines the advantages of Transformer framework and optimal truncated singular value decomposition (OT-SVD). TEAD comprises a multi-head transformer encoder for effective time series representation learning, a convolutional decoder for reconstruction, and a memory network for predictive analysis. OT-SVD is used to remove the noise of the data and provide reference values for the calculation of reconstruction loss and prediction loss, and the reconstruction loss and prediction loss are fused to calculate the anomaly score. Then, according to the anomaly score, the measured value of the sensor is judged to be normal or lot. Compared with the typical anomaly detection algorithms based on machine learning and deep learning, the proposed TEAD algorithm improves the detection performance and enhances the generalization ability of the model. On the SWAT dataset, the AUC value of TEAD algorithm is up to 0.916, while the AUC of the benchmarks algorithm is less than 0.85.

Key words: multi-sensor systems; anomaly detection; Transformer encoder; multi-head attention mechanism; truncated singular value decomposition
EEACC: 7220; 7230 **doi**: 10.3969/j.issn.1004-1699.2025.11.015

面向多传感器系统的基于 Transformer 编码器的异常检测算法*

李 岩[1], 倪笑宇[2], 陈 玉[3*]
(1.张家口职业技术学院机电工程学院, 河北 张家口 075000; 2.河北建筑工程学院机械工程学院, 河北 张家口 075000;
3.张家口开放大学教务处, 河北 张家口 075000)

摘 要: 多传感系统中的异常检测在工业监控、网络入侵检测等应用中起着至关重要作用。然而, 大量未标签的数据以及在时序数据中识别非异常情况给异常检测任务带来了巨大挑战。为此, 提出基于 Transformer 编码器的异常检测算法(Transformer-Encoder-based Anomaly Detection, TEAD)。TEAD 算法不依赖标签数据, 属于无监督学习, 它结合了 Transformer 框架和最优截断奇异值分解(Optimal Truncated Singular Value Decomposition, OT-SVD)的优点。TEAD 算法由用于有效时间序列表示的多头 Transformer 编码器、用于重构的卷积解码器和用于预测分析的预测模块构成。通过 OT-SVD 去除数据中的噪声, 为计算重构损失和预测损失提供参考值, 并对重构损失和预测损失进行融合, 进而计算异常分数。再依据异常分数判断传感器的测量值是否存在异常。与基于典型的机器学习和深度学习的异常检测算法相比, 所提出的 TEAD 算法提升了检测性能, 增强了模型的泛化能力。TEAD 算法在数据集 SWAT 上 AUC 值达到 0.916, 而基准算法的 AUC 不及 0.85。

关键词: 多传感器系统; 异常检测; Transformer 编码器; 多头注意力机制; 截断奇异值分解
中图分类号: TN929.5 **文献标识码**: A **文章编号**: 1004-1699(2025)11-2039-09

多传感器系统(Multi-Sensor Systems, MSS)中的异常检测在各领域都具有重要意义, 如工业监控[1]、网络入侵检测[2-3]、无线传感器网络[4]、医疗应用和自动驾驶领域。MSS 产生了复杂、高维的数据流, 这些数据流捕获了多个物理参数的时间演变信息。这些信息可用于检测数据流的异常或者也可用于对传感器的输出进行预测。

在工业应用中, 多传感器系统中的异常检测对

项目来源: 河北省高等学校科学技术研究项目(QN2021051); 河北省高等学校基本科研业务费项目(2024QNJS01)
收稿日期: 2024-11-08 **修改日期**: 2025-01-07

于诸如制造过程、发电厂和炼油厂等工业系统的健康或安全监测至关重要。检测异常可以帮助识别系统故障、预测设备故障,进而提高整体运营效率。类似地,在网络入侵检测中,异常检测能够识别异常的流量或者异常的用户行为。

海量的数据推动了机器学习方法的发展,催生了基于机器学习的异常检测算法。依据是否需要有标签的数据,这些算法可分为监督、半监督或无监督学习三类[5]。监督和半监督学习模式存在一些局限性:①只适用于有标签数据的应用场景;②只用于从训练数据中识别特征异常,可能无法识别未知的异常类型。

然而,动态的传感环境要求不断地使用新的带标签数据训练模型,才能检测传感环境的异常。但是,收集传感环境数据的速度不及传感环境中衍生异常或者攻击数据的速度,这就形成数据稀缺问题,最终可能导致模型过度拟合或者降低了模型的泛化能力。据此,无监督异常检测算法被广泛用于异常检测。它无需带标签数据,且能处理不平衡的数据问题,且具有检测未知的异常攻击,适应数据分布不断变化的非静态环境。

目前,无监督的异常检测方法可分为三类①依赖于经典的统计或机器学习的传统算法;②基于深度神经网络(Deep Neural Network,DNN)方法;③前两者的混合方法。

传统的方法有自回归模型[6]、不一致搜索和聚类算法[7]。这些方法的优点是简单、具有可解释性。然而,它们的性能往往受阻于高维、非线性的数据,导致它们不适用于复杂的 MSS 环境。

基于 DNN 的方法有递归神经网络(Recurrent Neural Networks,RNNs)[8-9],卷积神经网络(Convolutional Neural Networks, CNNs)[10-11],自动编码器[12],生成对抗网络[13-14],和图形神经网络。这些方法擅长捕捉复杂的非线性相互作用和时间相关性。然而,它们的主要缺点是需要大量的数据集和高计算资源。

混合方法将传统的机器学习与深度学习相结合,如具有变分自编码器的长短期记忆网络[15],多尺度卷积递归编码器-解码器网络,深度自编码高斯混合模型和反向训练的自编码器。由于顺序时序数据降低了训练 RNN 的速度,并且长期受到梯度消失问题挑战。研究人员提出 Transformer 模型。该模型能够通过多头自注意机制有效地处理上述问题,并且该模型支持并行序列处理,已成为多传感器时间序列顺序建模的最新技术。

为此,针对多传感器网络,提出基于 Transformer 编码器的异常检测算法(Transformer-Encoder-based Anomaly Detection, TEAD)。TEAD 算法先利用 Transformer 编码器捕获时空特征信息。Transformer 编码器采用了多头注意力机制,使模型更关注于重要特征。同时,采用卷积网络构建解码器,重构样本,并利用预测模块对传感器的测量值进行预测。考虑到输入数据的噪声问题,利用截断奇异值分解去除噪声,为计算重构损失和预测损失提供参考值。最终,将重构损失和预测损失进行融合,再计算异常分数,最后通过异常分数判断样本是否为攻击样本。

通过两个典型的数据集验证 TEAD 算法的性能。性能分析表明,提出的 TEAD 算法提升了检测性能,同时缩短了检测样本的时间,增强了 TEAD 算法的泛化能力。

1 问题表述及整体框架

1.1 问题表述

假定系统中有 S 个传感节点。令 $x_k[n]$ 表示在离散时刻 n 第 k 个传感节点的测量值。因此,在时刻 n 所获取的测量信息可表示成一个测量数据矢量 $\boldsymbol{x}[n]$:

$$\boldsymbol{x}[n]=(x_1[n],x_2[n],\cdots,x_S[n])^{\mathrm{T}} \tag{1}$$

因此,在 N 个离散时刻内,整个多传感系统的测量信息可表示成测量数据矩阵 $\boldsymbol{X}$:

$$\boldsymbol{X}=(\boldsymbol{x}[1],\boldsymbol{x}[2],\cdots,\boldsymbol{x}[N]) \tag{2}$$

假定训练数据集为 $\boldsymbol{X}_{\mathrm{train}}$,其表示正常条件下的 MSS 系统的行为。训练过程的目的在创建准确的表示 $G(\cdot)$,以捕获系统的正常行为。令 $\boldsymbol{X}_{\mathrm{test}}$ 表示测试数据集,其包含正常和异常情况的数据。相应地,令 $\boldsymbol{y}$ 表示 $\boldsymbol{X}_{\mathrm{test}}$ 所对应的标签矢量:

$$\boldsymbol{y}=(y_1,y_2,\cdots,y_m,\cdots,y_M)^{\mathrm{T}} \tag{3}$$

式中:若 $y_m=1$,则表示在离散时刻 m 存在攻击,反之表示不存在攻击。

训练模型的目的在于构建一个表示 $F(\cdot)$,使得 $F(\cdot)$ 对 $\boldsymbol{X}_{\mathrm{test}}$ 的估计值 $\hat{\boldsymbol{y}}=(\boldsymbol{X}_{\mathrm{test}})$ 更逼近于 $\boldsymbol{y}$。

1.2 整体框架

TEAD 模型采用一种基于预测和重构的新型模块化双重架构,进而解决多传感系统中无监督异常检测的挑战。整个框架如图 1 所示,主要由重构网络和潜在预测网络构成。这两个网络有一个共同的模块:Transformer 编码器。该编码器用于捕获时空特征信息。卷积解码器用于执行重构任务;预测模块用于执行预测任务。检测模块通过融合预测损失和重构损失,完成对输入数据的异常检测。截断奇异值分解用于去除原始数据中的噪声。

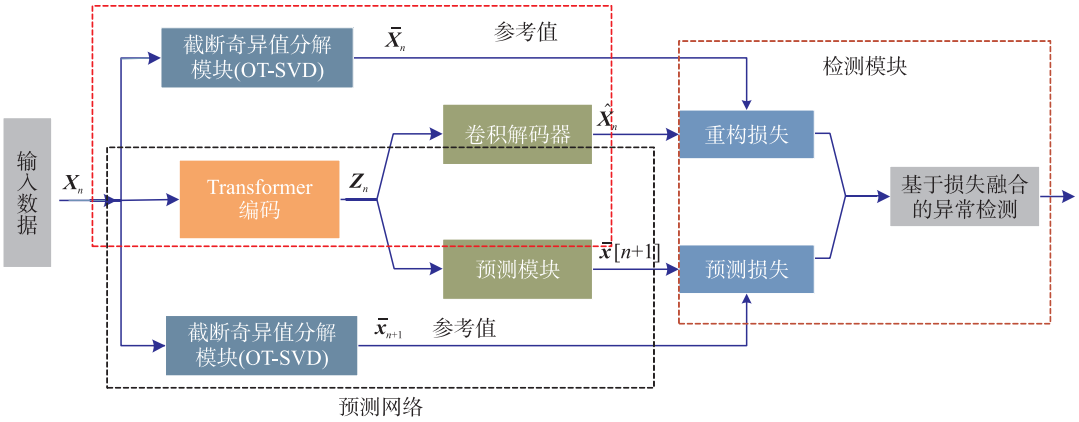

图 1　TEAD 算法的整体框架

2　TEAD 模型

2.1　基于滑动窗口的数据流生成

由于 TEAD 模型处理的数据是数据流,需通过滑动窗口将数据转换成数据流格式。令 $\boldsymbol{X}_n$ 表示系统在时刻 n 的测量样本,其是经滑动窗口处理后的序列数据流:

$$\boldsymbol{X}_n=(\boldsymbol{x}[n],\boldsymbol{x}[n-1],\cdots,\boldsymbol{x}[n-L+1]) \tag{4}$$

式中:L 表示窗口尺寸。图 2 给出输入的序列数据流的示例。

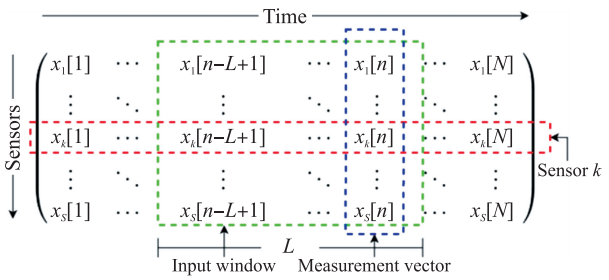

图 2　序列数据流示例

2.2　Transformer 编码器

Transformer 编码器可捕获多传感器数据中复杂的时空模式,在 TEAD 算法中起着至关重要的作用。Transformer 编码器由 L_T 个 Transformer 编码单元构成,如图 3 所示。

每个 Transformer 编码单元由多头自注意模块、位置前馈网络和带有残差连接的归一化层组成。与递归网络不同,Transformer 网络以并行方式处理输入序列数据,导致序列信息丢失。据此,需通过位置编码矩阵 $\boldsymbol{P}$ 对位置进行编码,以避免序列信息的丢失。

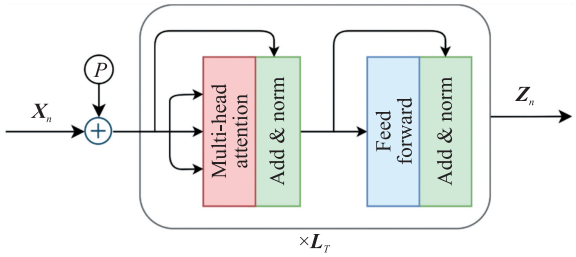

图 3　Transformer 编码器结构

2.2.1　多头注意力机制

自注意力机制(Self-Attention)通过考虑序列中位置,捕捉序列内的长距离依赖关系。它在处理序列数据时主要涉及到查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$。用 $\boldsymbol{Q}_i$、$\boldsymbol{K}_i$ 和 $\boldsymbol{V}_i$ 分别表示第 i 个注意头的查询矩阵、键矩阵和值矩阵 $\boldsymbol{V}$,它们的定义如式(5)所示:

$$\begin{cases}\boldsymbol{Q}_i=(\boldsymbol{X}_n+\boldsymbol{P})\boldsymbol{W}_{Q,i}\\ \boldsymbol{K}_i=(\boldsymbol{X}_n+\boldsymbol{P})\boldsymbol{W}_{K,i}\\ \boldsymbol{V}_i=(\boldsymbol{X}_n+\boldsymbol{P})\boldsymbol{W}_{V,i}\end{cases} \tag{5}$$

式中:$\boldsymbol{W}_{Q,i}\in\mathbf{R}^{S\times(S/h)}$、$\boldsymbol{W}_{K,i}\in\mathbf{R}^{S\times(S/h)}$ 和 $\boldsymbol{W}_{V,i}\in\mathbf{R}^{S\times S}$ 分别表示学习的权重系数;h 表示注意力的层数。

令 $\boldsymbol{C}$ 表示上下文矩阵,其表示自注意力单元的输出,其可表示为:

$$\boldsymbol{C}=\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{6}$$

式中:d_k 表示数据维度。

图 4 给出了多头自注意力机制的网络结构模型。

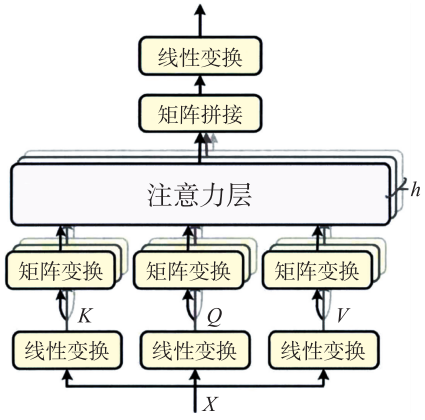

图 4　多头自注意力机制的网络结构图

如图 4 所示,先通过 $\boldsymbol{X}$ 与 $\boldsymbol{W}_{Q,i}$、$\boldsymbol{W}_{K,i}$ 和 $\boldsymbol{W}_{V,i}$ 相乘,得到矩阵 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$,再将它们分别转换成 h 个矩阵:

$$\langle \boldsymbol{Q}_1,\boldsymbol{Q}_2,\cdots,\boldsymbol{Q}_h;\boldsymbol{K}_1,\boldsymbol{K}_2,\cdots,\boldsymbol{K}_h;\boldsymbol{V}_1,\boldsymbol{V}_2,\cdots,\boldsymbol{V}_h\rangle \tag{7}$$

这 h 组中的每一组 $\boldsymbol{Q}_i$、$\boldsymbol{K}_i$ 和 $\boldsymbol{V}_i$ 都对应一个注意力层,其中 $i=1,2,\cdots,h$。经注意力机制处理后,将结果进行拼接。最后,将拼接后的结果作为线性变换的输入,最终得到自注意力机制的输出:

$$\mathrm{MHV}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\boldsymbol{W}\mathrm{Concat}(\mathrm{head}_1,\cdots,\mathrm{head}_h) \tag{8}$$

式中:$\boldsymbol{W}$ 是权重矩阵;Concat(·)表示拼接操作;head_i 表示第 i 个头的输出,$i=1,2,\cdots,h$。

$$\mathrm{head}_i=\mathrm{Attention}(\boldsymbol{Q}\boldsymbol{W}_{Q,i},\boldsymbol{K}\boldsymbol{W}_{K,i},\boldsymbol{V}\boldsymbol{W}_{V,i})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{W}_{Q,i}(\boldsymbol{K}\boldsymbol{W}_{K,i})^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}\boldsymbol{W}_{V,i} \tag{9}$$

2.2.2 Transformer 编码器的输出

将多头注意力的输出与输入数据相加,再经归一化层处理,如图 3 所示。然后,由位置前馈神经网络(Position-wise Feed-forward Neural Network,FFNN)处理,进而捕获复杂的非线性行为,最后再与输入相加,并经归一化处理。归一化层,对输入数据进行归一化,减少协变量移位问题,有助于提升学习过程中的稳定性。

最终,将 Transformer 编码器的输出表示为:

$$\boldsymbol{Z}_n=I^{L_T}(\boldsymbol{X}_n+\boldsymbol{P}) \tag{10}$$

式中:$\boldsymbol{Z}_n$ 表示 Transformer 编码器的输出;$I(\cdot)$ 表示 Transformer 编码单元的操作;$I^{L_T}(\cdot)$表示 Transformer 编码器的操作(其由 L_T 个 Transformer 编码单元构成)。

2.3 卷积解码器

通过卷积解码器,对 Transformer 编码器的输出 $\boldsymbol{Z}_n$ 进行重构。解码器由一维卷积网络构成,并由它执行重构任务。令 $C(\cdot)$表示一维卷积操作:

$$\hat{\boldsymbol{X}}_n=C(\boldsymbol{Z}_n) \tag{11}$$

式中:$\hat{\boldsymbol{X}}_n$ 表示卷积解码器的输出,即对 $\boldsymbol{Z}_n$ 重构后的数据,进而解码器就可利用重构误差去检测多传感器测量中的异常。

2.4 预测模块

预测模块的作用在于:基于 Transformer 编码器的输出 $\boldsymbol{Z}_n$,预测传感器未来的测量值($n+1$ 时刻的值)。它由基于 Sigmoid 激活函数的前馈网络构成。具体而言,基于 $\boldsymbol{Z}_n$,通过线性变换预测 $n+1$ 时刻的传感器测量矢量:

$$\breve{\boldsymbol{x}}[n+1]=F(\boldsymbol{Z}_n) \tag{12}$$

式中:$F(\cdot)$表示预测模块的操作;$\breve{\boldsymbol{x}}[n+1]$表示预测 $n+1$ 时刻的传感器测量矢量。

预测模块的关键在于它能够同时检测点异常和序列异常。这种双重检测能力提高了检测性能。当与重构解码器配对时,预测模块为多传感器数据中的无监督异常检测提供了鲁棒的检测机制。

2.5 截断奇异值分解模块

为了消除噪声对异常检测的影响,TEAD 算法利用截断奇异值分解(OT-SVD)消除训练阶段数据中的噪声。OT-SVD 模块通过在噪声数据中鉴别真实异常值,进而提升检测准确率。OT-SVD 依赖于矩阵的奇异值:

$$\boldsymbol{X}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}=\sum_{i=1}^{\min(S,L)}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}} \tag{13}$$

式中:$\boldsymbol{U}\in\mathbf{R}^{S\times S}$ 和 $\boldsymbol{V}\in\mathbf{R}^{L\times L}$ 均是正交矩阵,即 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}_S$、$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{I}_L$ 和 $\boldsymbol{\Sigma}\in\mathbf{R}^{S\times L}$是非负对角矩阵。

令 $\boldsymbol{u}_i$ 表示矩阵 $\boldsymbol{U}$ 的列,它们是矩阵 $\boldsymbol{X}$ 的左奇向量,即是 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 的特征向量。类似地,$\boldsymbol{v}_i^{\mathrm{T}}$ 表示矩阵 $\boldsymbol{V}^{\mathrm{T}}$ 的列,它们是矩阵 $\boldsymbol{X}$ 的右奇向量,对应是 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 的特征向量。用 σ_i 表示 $\boldsymbol{\Sigma}$ 的对角元素。

多传感系统,通常需要观察来自不同传感器测量值之间的依赖性。因此,假定由传感器捕获的信号为低阶结构是合理的。据此,采用低秩近似是一种非常有吸引力的数据表示方法。具体而言,采用截断奇异值(Truncated Singular Value Decomposition,TSVD)估计法。TSVD 估计法是一种通过截断奇异值来减少噪声影响的数据处理方法。TSVD 估计法通过仅保留原始矩阵 $\boldsymbol{X}$ 的前 r 个奇异值以及其相关奇异向量来最小化 Frobenius 范数:

$$\widetilde{\boldsymbol{X}}_{(r)}=\underset{\hat{\boldsymbol{X}};\mathrm{rand}(\hat{\boldsymbol{X}})\leqslant r}{\arg\min}\|\boldsymbol{X}-\hat{\boldsymbol{X}}\|_F^2=\sum_{i=1}^{r}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}=\widetilde{\boldsymbol{U}}_{(r)}\widetilde{\boldsymbol{\Sigma}}_{(r)}\widetilde{\boldsymbol{V}}_{(r)}^{\mathrm{T}} \tag{14}$$

式中:$\widetilde{\boldsymbol{U}}_{(r)}$ 和 $\widetilde{\boldsymbol{V}}_{(r)}^{\mathrm{T}}$ 分别表示 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的第 r 列;$\widetilde{\boldsymbol{\Sigma}}_{(r)}$ 表示 $\boldsymbol{\Sigma}$ 中前 $r\times r$ 子块。

由图 1 可知,OT-SVD 模块的作用在于为卷积编码器和预测模块提供无噪声的参考值。具体而言,令 $\overline{\boldsymbol{X}}_n$ 表示 OT-SVD 模块为卷积编码器提供的参考值,其是经 OT-SVD 模块处理后 OT-SVD 模块的输出值,输入值是 $\boldsymbol{Z}_n$。令 $\bar{\boldsymbol{x}}[n+1]$表示 OT-SVD 模块为预测模型提供的参考值。

2.6 模型训练

TEAD 算法涉及两处损失:重构损失和预测损失,如图 1 所示。通过将这两处损失进行融合,构成一个总体的损失函数,并以最小化该损失函数为训练目标,进而提升模型的检测能力。

具体而言,所谓的重构损失,就是 OT-SVD 模块

的输出矩阵 $\bar{\boldsymbol{X}}_n$ 与卷积编码器的输出矩阵 $\hat{\boldsymbol{X}}_n$ 间的均方误差(Mean Squared Error,MSE),其中 $\bar{\boldsymbol{X}}_n=(\bar{\boldsymbol{x}}[n],\bar{\boldsymbol{x}}[n+1],\cdots,\bar{\boldsymbol{x}}[n+L-1])$,重构序列 $\hat{\boldsymbol{X}}_n=(\hat{\boldsymbol{x}}[n],\hat{\boldsymbol{x}}[n+1],\cdots,\hat{\boldsymbol{x}}[n+L-1])$。据此,单个序列的重构损失可为表述为:

$$L_{\text{rec}}=\frac{1}{S}\frac{1}{L}\sum_{k=1}^{S}\sum_{\ell=0}^{L-1}[\bar{x}_k[n-\ell]-\hat{x}_k[n-\ell]]^2 \quad (15)$$

预测模块采用 Transformer 编码器和预测网络以自回归方式预测每个滑动窗口的未来值。预测损失就是 OT-SVD 模块的输出矩阵 $\bar{\boldsymbol{x}}_{n+1}$ 与卷积编码器的输出矩阵 $\breve{\boldsymbol{x}}_{n+1}$ 间的均方误差 MSE:

$$L_{\text{pred}}=\frac{1}{S}\sum_{k=0}^{S}[\bar{x}_k[n+1]-\hat{x}_k[n+1]]^2 \quad (16)$$

对再 L_{rec} 和 L_{pred} 进行融合,构建总体的损失:

$$L=\alpha L_{\text{rec}}+\beta L_{\text{pred}} \quad (17)$$

式中:α 和 β 为权重系数,且 $\alpha+\beta=1$。

TEAD 算法以最小化式(17)损失函数,训练模型,获取最优的异常检测模型。

2.7　异常检测

TEAD 算法将预测误差和重构误差作为异常分数,再将异常分数与阈值进行比较,进而检测是否存在异常。通过 ℓ_2-norm 范数或者欧氏距离计算误差。依据式(18)计算异常分数 s_m:

$$s_m=\alpha_s\|\boldsymbol{x}[m]-\hat{\boldsymbol{x}}[m]\|_2+\beta_s\|\boldsymbol{x}[m]-\breve{\boldsymbol{x}}[m]\|_2 \quad (18)$$

式中:α_s 和 β_s 分别表示在测试阶段的重构误差和预测误差的权重系数。

若 s_m 大于 λ^*,则输入序列为异常($\hat{y}_m=1$),反之无异常($\hat{y}_m=0$):

$$\hat{y}_m=\begin{cases}1 & s_m>\lambda^*\\ 0 & s_m\leqslant\lambda^*\end{cases} \quad (19)$$

式中:λ^* 表示阈值。

利用训练阶段的异常分数 s_m 的均值和方差,计算阈值 λ^*:

$$\lambda^*=\hat{\mu}_s+3\hat{\sigma}_s \quad (20)$$

式中:$\hat{\mu}_s=(1/N)\sum_{n=1}^{N}s_m$ 表示异常分数 s_m 的均值;$\hat{\sigma}_s=(1/N)\sum_{n=1}^{N}(s_n-\hat{\mu}_s)^2$ 表示异常分数 s_m 的方差。

算法 1 给出了检测异常的具体过程。算法 1 的输入为:测试数据集 $\boldsymbol{X}_{\text{test}}=(\boldsymbol{x}[1],\boldsymbol{x}[2],\cdots,\boldsymbol{x}[M])$,其包含正常和异常数据;滑动窗口尺寸 L;测试数据集的真实标签 $\boldsymbol{y}=(y_1,y_2,\cdots,y_M)^{\mathrm{T}}$;阈值 λ^*。算法 1 的输出为:对标签的估计值:$\hat{\boldsymbol{y}}=(\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_M)^{\mathrm{T}}$。

对于任意一个样本 m,先通过卷积编码器对其进行重构,生成重构样本 $\hat{\boldsymbol{x}}_m$,并通过预测模块进行预测,生成预测样本 $\breve{\boldsymbol{x}}_m$。再依据式(18)计算异常分数。

最后,将异常分数 s_m 与阈值 λ^* 进行比较,得到标签的估计值。

算法 1　异常检测过程

输入:$\boldsymbol{X}_{\text{test}}=(\boldsymbol{x}[1],\boldsymbol{x}[2],\cdots,\boldsymbol{x}[M])$;窗口尺寸 L;
　　$\boldsymbol{y}=(y_1,y_2,\cdots,y_M)^{\mathrm{T}}$;阈值 λ^*;
输出:$\hat{\boldsymbol{y}}=(\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_M)^{\mathrm{T}}$;
1. For $m\leftarrow1$ to M do
2. 　$\hat{\boldsymbol{x}}_m\leftarrow C(I(\boldsymbol{X}_n))$,$\breve{\boldsymbol{x}}_m\leftarrow F(I(\boldsymbol{X}_n))$;
3. 　依据式(18)计算异常分数 s_m;
4. 　if $s_m>\lambda^*$ then $y_m=1$
5. 　　else $y_m=0$
6. 　end if
7. end for

3　性能分析

利用 PyTorch 和 TensorFlow 深度学习框架模型进行模型训练和评估。此外,利用 scikit-learn 机器学习库中的函数对数据进行预处理。

3.1　选用的数据集

选用两个公开的真实世界的多传感器数据集评估所提出的 TEAD 模型的性能。

①数据集 SWAT[16-17]。SWAT 模拟了一个水处理厂的工业控制系统,并记录了在该系统中发生的各种操作和事件,其广泛用评估物联网系统的安全以及可能存在的攻击和漏洞。它包括 11 天的连续运行数据(正常运行 7 天,攻击为 4 天)。

②数据集 WADI[18]。WADI 是 SWAT 的扩展版,是一个广泛使用的网络入侵检测数据集。它包含 16 天的数据(14 天的正常运行和 2 天的攻击场景)。

表 1 给出了 SWAT 和 WADI 两个数据集的训练样本数和测试样本数。

表 1　数据集 SWAT 和 WADI 的数据分布

数据集	训练样本数	测试样本数	攻击样本占比/%
SWAT	495 000	449 919	12.14
WADI	1 209 601	172 801	5.99

3.2　基准算法

为了评估所提出的方法的性能,选择了以下最先进、常规和深度学习异常检测方法作为基准模型。

①孤立森林(Isolation Forest,IF)[19]。IF 是基

于决策树的无监督异常检测算法。利用开源的 PyOD Python 包实现 IF。IF 采用 100 棵树集合作为其估计器,其中每棵树使用单个特征在单个节点处执行分裂。

②多层感知器自动编码器(Multilayer Perception Autoencoder,MLP-AE)[12]。MLP-AE 是一个深度学习算法,它使用前馈编码器-解码器神经网络重构原始样本数据,并通过重构误差识别样本是否为攻击样本。在实验中,MLP-AE 由三层编码器和三层解码器组成。

③无监督的异常检测(Unsupervised Anomaly Detection,USAD)[20]。USAD 是一种基于逆向训练的自编码器的方法。USAD 方法使用默认超参数实现。该模型架构包括一个编码器和两个解码器,其中每个模块由三个线性层组成,在三个线性层之间具有整流线性单元(ReLU)激活函数,并且最后一层采用 Sigmoid 激活函数。

④基于生成对抗网络的多变量异常检测(MAD-GAN)[21]。MAD-GAN 是一种无监督方法,其利用 GAN 学习深层正态数据分布,并基于真实数据与生成数据之间的差异检测是否存在异常。MAD-GAN 利用长短期记忆网络(Long Short-Term Memory,LSTM)作为生成器和鉴别器。

3.3 评估指标

结合标签数据,本文将异常检测问题视为二分类任务。利用 F1、接受者操作特征(Receiver Operating Characteristic)曲线下面积(AUC)和精确-召回(Precision-Recall,PR)曲线下面积(PRC)三个指标综合评估 TEAD 模型的性能。

F1 的定义如式(21)所示。F1 指标融合了 Precision 和 Recall,它兼顾了模型的精确率和召回率,可以看作是精确率和召回率的调和平均数。因此,选择 F1 没有选 Precision 和 Recall 作为性能指标。

$$\mathrm{F1}=2\times\frac{\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{21}$$

式中:Precision 和 Recall 分别表示精确率和准确率,它们的定义如式(22)所示:

$$\begin{cases}\mathrm{Precision}=\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\times 100\% \\ \mathrm{Recall}=\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\times 100\%\end{cases} \tag{22}$$

式中:TP 表示模型正确预测为正的样本数;FP 表示模型错误预测为正的样本数;FN 表示模型错误预测为负的样本数。

3.4 关键参数对模型性能的影响

TEAD 算法主要涉及到三个关键感参数:滑动窗口尺寸 L、Transformer 编码器中包含的 Transformer 编码单元的个数 L_T 和多头注意力模块中多头数 h。这三个参数的默认取值分别为:$L=200$,$L_T=1$ 和 $h=2$。

首先,图 5(a)给出了 F1、AUC 和 PRC 随滑动窗口尺寸 L 的变化情况,其中 L 从 50 至 250 变化,步长为 50。$L_T=1$,$h=2$。从图 5(a)可知,随着 L 值的增加,三个指标(F1、AUC 和 PRC)缓慢增加,呈增加趋势。这说明增加 L 可以提升检测性能。但是 L 值的增加,加重了训练模型的运算量。

图 5(b)给出了 F1、AUC 和 PRC 随 L_T 的变化情况,其中 L_T 从 1 至 4 变化,步长为 1。$L=200$,$h=2$。从图 5(b)可知,Transformer 编码单元的个数 L_T 与三个指标(F1、AUC 和 PRC)并非呈单调递增关系。这表明更复杂的模型不一定能提供更优的检测性能,很可能会导致过拟合。

最后,图 5(c)给出了 F1、AUC 和 PRC 随滑动窗口尺寸 h 的变化情况,其中 $h\in\{2,4,8,16,32\}$。$L=200$,$L_T=1$。从图 5(c)可知,F1、AUC 和 PRC 并没有随 h 的增加而呈有规律的变化,即并非单调关系。

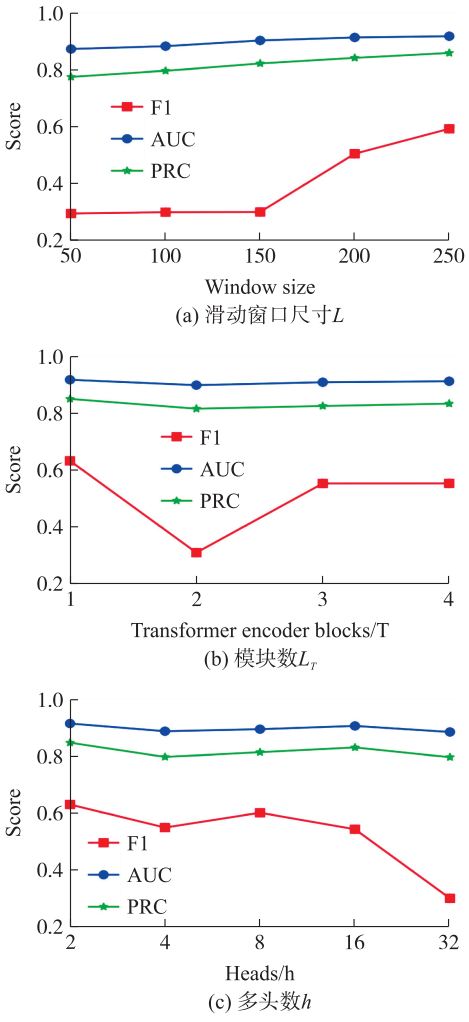

图 5 F1、AUC 和 PRC 随着 L、L_T 和 h 的变化情况

3.5　OT-SVD 模块对模型性能的影响

TEAD 算法利用 OT-SVD 模块去除数据中的噪声,为计算重构损失和预测操作提供无噪声的参考数据。为了验证 OT-SVD 模块是否对模型性能有积极的作用,将未采用 OT-SVD 模块的 TEAD 算法(标记为 TEAD-U)与 TEAD 算法进行比较。

图 6 给出了 TEAD-U 和 TEAD 算法的 F1、AUC 和 PRC 三个指标的性能。从图可知,相比于 TEAD-U 算法,TEAD 算法均不同程度地提升了 F1、AUC 和 PRC。这表明,通过 OT-SVD 模块去除噪声,可提升模型的检测性能。以 F1 为例,TEAD-U 算法的 F1 值约 0.582,而 TEAD 算法的 F1 值达到 0.653,提升了约 12%。

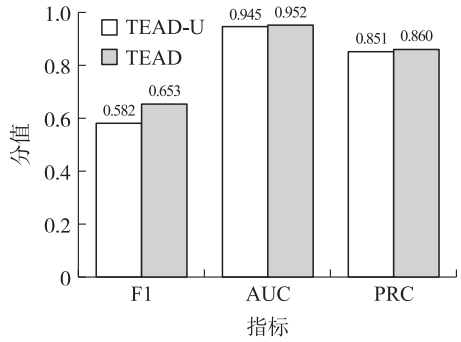

图 6　OT-SVD 模块对模型性能的影响

3.6　与基准算法的比较

本节将 TEAD 算法与 IF、MLP-AE、USAD、MAD-GAN 四个基准算法进行比较,分析它们的 F1、AUC 和 PRC 性能。

图 7 给出了基于数据集 SWAT 上的 TEAD 和基准模型的异常检测性能。从图可知,相比于 IF、MLP-AE、USAD、MAD-GAN 算法,提出的 TEAD 模型有效地提升了检测性能。从 F1 指标来看,IF、MLP-AE、USAD 和 MAD-GAN 算法的 IF 值均不到 0.4,而提出的 TEAD 模型的 IF 达到 0.629,提升了约 50%。

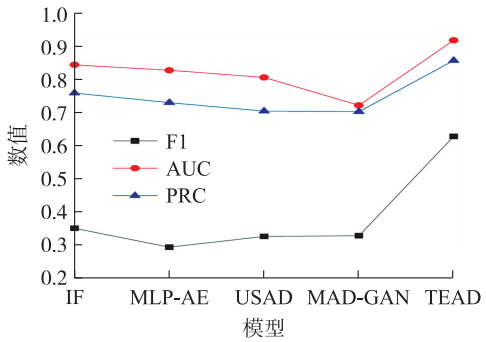

图 7　基于数据集 SWAT 上的 TEAD 和基准算法的异常检测性能

从 AUC 指标来看,TEAD 算法的性能仍优于 IF、MLP-AE、USAD 和 MAD-GAN 算法。TEAD 算法的 AUC 达到 0.916,而 IF、MLP-AE、USAD 和 MAD-GAN 算法的 AUC 分别为 0.842、0.826、0.804 和 0.721,均不及 0.85。此外,TEAD 算法在 PRC 方面的性能仍优于 IF、MLP-AE、USAD 和 MAD-GAN 算法。

上述数据表明,TEAD 算法通过 Transformer 编码器对输入样本进行编码,并通过卷积解码器解码,可有效地检测传感器的异常值。

类似地,图 8 给出了基于数据集 WADI 上的 TEAD 和基准算法的异常检测性能。从图可知,即使在数据集 WADI 上,TEAD 算法的在 F1、AUC 和 PRC 三方面性能均优于基准算法。

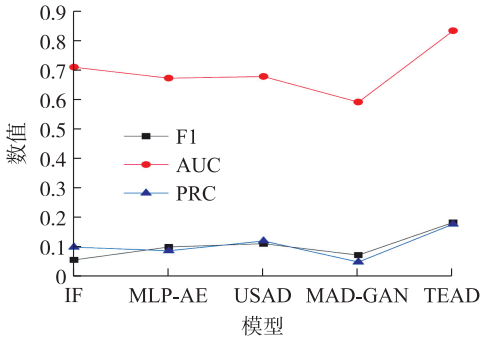

图 8　基于数据集 SWAT 上的 TEAD 和基准算法的异常检测性能

此外,对比图 7 和图 8 不难发现,TEAD 算法和 4 个基准算法在数据集 SWAT 上的检测性能优于在数据集 WADI 上的性能。可能的原因在于:数据集 WADI 的样本数较多,且分布不均衡,这增加了执行异常检测任务的难度。

3.7　复杂度分析

TEAD 算法的总体复杂度由多个模块决定。由于 OT-SVD 只在预训练阶段完成,其不考虑到算法的复杂度中。,TEAD 算法中最复杂的计算来自 Transformer 编码器。

具体而言,每个多头注意力层的计算复杂度为 $O(L^2\times S)$。Transformer 中前馈网络增加了额外的复杂度,其复杂度为 $O(L\times S^2)$。总的计算复杂度与 L_T 值成正比。解码器采用一维卷积层,其专注于重构输入序列。卷积层的复杂度为 $O(k\times L\times S^2)$,其中 k 表示内核大小。

预测模型是前馈神经网络,它的复杂度主要取决于层数,每层的神经元数以及每个神经元的操作。复杂度等于当前神经元数 n_ℓ 和前一层神经元数 $n_{\ell-1}$ 的乘积,加上当前层中每个神经元的偏置项,即 $O(n_{\ell-1}\times n_{\ell-1}+n_{\ell-1})$。

图 9 给出了 IF、MLP-AE、USAD、MAD-GAN 和 TEAD 算法的平均训练时间和测试时间。从图可知,IF、MLP-AE、USAD 和 MAD-GAN 算法的训练时间分别为 0.3591 s、0.2847 s、81.0371 s 和 80.49 s。而 TEAD 算法的训练时间为 24.067 s。由于 IF 和 MLP-AE 算法是传统的机器学习,并没有采用深度神经网络,它们的训练时间较短。

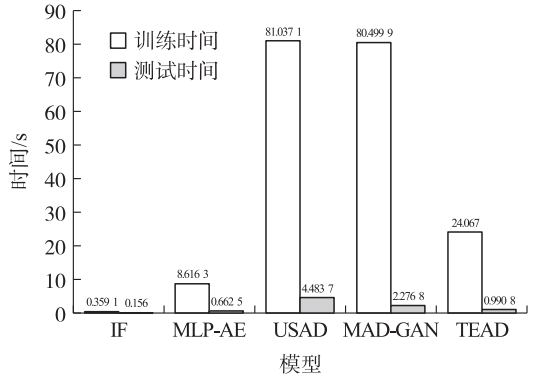

图 9 TEAD 算法的训练时间和测试时间

USAD、MAD-GAN 和 TEAD 算法同属于深度神经网络。但是 TEAD 算法的训练时间远低于 USAD 和 MAD-GAN 算法的训练时间。此外,TEAD 算法的测试时间也低于 USAD 和 MAD-GAN 的测试时间。

4 结束语

本文针对多传感器系统的异常检测问题,提出了一个基于 Transformer 编码器的异常检测算法 TEAD。TEAD 算法由重构网络和潜在预测网络两部分组成。此外,TEAD 算法利用 OT-SVD 解决了传感器测量值的噪声问题。同时,通过多头注意力机制解决传感器测量值之间的时空相关性。此外,TEAD 算法先通过融合重构损失和预测损失计算异常分数,再基于异常分数判断是否有异常。仿真结果表明,同传统的机器学习和深度学习的异常算法相比,TEAD 算法提高了 F1 和 AUC 值,提升了模型检测异常的效率。在 SWAT 数据集上,TEAD 算法的 AUC 达到 0.916,而 IF、MLP-AE、USAD 和 MAD-GAN 算法的 AUC 分别为 0.842、0.826、0.804 和 0.721。此外,TEAD 算法缩短了检测时间,它检测单个样本的平均时间为 0.990 8 s,而 MAD-GAN 算法的检测时间达到 2.276 8 s。

参考文献:

[1] Nizam H,Zafar S,Lv Z,et al. Real-Time Deep Anomaly Detection Framework for Multivariate Time-Series Data in Industrial IoT[J]. IEEE Sensor Journal,2022,22(23):22836-22849.

[2] 刘奇旭,肖聚鑫,谭耀康,等. 工业互联网流量分析技术综述[J]. 通信学报,2024,45(8):221-237.

[3] 张杭生,刘吉强,梁杰,等. 基于博弈论的入侵检测与响应优化综述[J]. 信息安全学报,2024,9(4):163-179.

[4] 丁华峰,胡学龙. 基于卡尔曼滤波的无线传感器网络信息并行传输延迟补偿算法[J]. 传感技术学报,2024,37(10):1814-1819.

[5] 刘刚,叶立新,陈麒玉,等. 基于多传感器信息融合的城市边坡监测数据异常事件检测[J]. 地质科技通报,2022,41(2):13-25.

[6] Melnyk I,Matthews B,Valizadegan H,et al. Vector Autoregressive Model-Based Anomaly Detection in Aviation Systems[J]. Journal Aerospace Information Systems,2016,13(4):161-173.

[7] Nie F,Li Z,Wang R,et al. An Effective and Efficient Algorithm for *K*-Means Clustering with New Formulation[J]. IEEE Transaction Knowledge Data Engineer,2023,35(4):3433-3443.

[8] 刘业峰,王帅,刘晶晶,等. 基于 TCN-SA 和 Bi-GRU 的轴承剩余寿命预测[J/OL]. 计算机集成制造系统,1-11(2024-08-26)[2024-11-17]. https://doi.org/10.13196/j.cims.2024.0260.

[9] 任利强,贾舒宜,王海鹏,等. 基于深度学习的时间序列分类研究综述[J]. 电子与信息学报,2024,46(8):3094-3116.

[10] 朱仔旭,施成华,陈海勇,等. 考虑 IPSO 优化的 CNN-LSTM 盾构掘进参数混合预测模型[J]. 铁道科学与工程学报,2025,22(6):2690-2702.

[11] 乔卉卉,赵二贤,郝如江,等. 基于注意力机制与多源信息融合的变工况轴承故障诊断[J]. 仪器仪表学报,2024,45(9):120-130.

[12] 杜谦,谢卫莹. 对抗性自动编码网络在高光谱异常检测中的应用[J]. 测绘学报,2023,52(7):1105-1114.

[13] 唐伦,赵禹辰,薛呈呈,等. 一种基于时间序列分解和时空信息提取的云服务器异常检测模型[J]. 电子与信息学报,2024,46(6):2638-2646.

[14] 陈鸿鑫,马天霆,周阳,等. 基于 CNN-GAN 数据增强网络的电厂锅炉管道温度压力及健康状态预测[J]. 电子器件,2023,46(6):1593-1600.

[15] Lin S Y,Clark R,Birke R,et al. Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model[C]//2020 IEEE International Conference on Acoustics,Speech and Signal ing,Barcelona,Spain,2020:4322-4326.

[16] Goh J,Adepu S,Junejo K N,et al. A Dataset to Support Research in the Design of Secure Water Treatment Systems[C]//11th International Conference on Critical Information Infrastructures Security,,Paris,France,2017:88-99.

[17] Mathur A P,Tippenhauer N O. SWAT:A Water Treatment Tested for Research and Training on ICS Security[C]// 2016 International Workshop on Cyber-physical Systems for Smart Water Network,Vienna,Austria,2016:31-36.

[18] Ahmed C M,Palleti V R,Mathur A P. WADI:A Water Distribution Tested for Research in the Design of Secure Cyber Physical Systems[C]//3rd International Workshop on Cyber-Physical Systems for Smart Water Network, Pittsburgh,PA,USA,2017:25-28.

[19] Liu F T,Ting K M,Zhou Z H. Isolation-Based Anomaly Detection[J]. ACM Transactions Knowledge Discovery from Data,2012,6(1):3.

[20] Audibert J,Michiardi P,Guyard F,et al. USAD:Unsupervised Anomaly Detection on Multi-Variate Time Series [C]//26th ACM SIGKDD International Conference Knowledge Discovery Data Ming,Online,2020:3395-3404.

[21] Zhao H,Wang Y J,Duan J Y,et al. Multivariate Time-Series Anomaly Detection via Graph Attention Network [C]//2020 IEEE International Conference on Data Mining,Sorrento,Italy,2020:841-850.

李　岩(1979—),男,汉族,吉林松原人,工程硕士,副教授,研究方向:电子与通信;

倪笑宇(1984—),男,汉族,河北张家口人,硕士,教授,研究方向:计算机视觉;机电一体化;

陈　玉(1978—),女,汉族,吉林白山人,工学学士,副教授,主要研究方向为计算机科学与技术。