计算机科学
COMPUTER SCIENCE

双编码半监督异常检测模型

李辉, 李文根, 关佶红

引用本文

李辉, 李文根, 关佶红.双编码半监督异常检测模型[J]. 计算机科学, 2023, 50(7): 53-59.

LI Hui, LI Wengen, GUAN Jihong. Dually Encoded Semi-supervised Anomaly Detection[J]. Computer Science, 2023, 50(7): 53-59.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220900027

双编码半监督异常检测模型

李 辉 李文根 关佶红
同济大学电子与信息工程学院 上海 201804
(2230760@tongji.edu.cn)

摘 要 异常检测是机器学习领域广泛研究的一个热点问题,对于工业生产、食品安全、疾病监测等都具有重要作用。当前最新的异常检测方法多基于少量可用的有标记样本和大量无标记样本联合训练半监督检测模型。然而,现有的半监督异常检测模型多采用深度学习框架,在低维数据集上由于缺少足够多的特征信息,难以学习到准确的数据边界,检测性能不佳。针对该问题,提出了双编码半监督异常检测模型(Dually Encoded Semi-supervised Anomaly Detection,DE-SAD),充分利用可获得的少部分有标记数据结合大量无标记数据进行半监督学习,通过双编码阶段约束模型学习更准确的正常数据隐含流形分布,有效拉大了正常数据和异常数据的差距。DE-SAD在来自不同领域的多个异常检测数据集上都表现出优越的异常检测性能,在低维数据上的检测性能尤为突出,其AUROC指标相比当前最优的异常检测方法最高提升了4.6%。

关键词:异常检测;半监督学习;自编码器;低维数据
中图法分类号 TP391

Dually Encoded Semi-supervised Anomaly Detection

LI Hui,LI Wengen and GUAN Jihong
College of Electronic and Information Engineering,Tongji University,Shanghai 201804,China

Abstract Anomaly detection is a hot topic that has been widely studied in the field of machine learning and plays an important role in industrial production,food safety,disease monitoring,etc. The latest anomaly detection methods mostly jointly train semi-supervised detection models based on a small number of available labeled samples and many unlabeled samples. However,these existing semi-supervised anomaly detection models mostly use deep learning frameworks. Due to the lack of enough feature information on low-dimensional data sets,it is difficult to learn accurate data boundaries,resulting in insufficient detection performance. To solve this problem,a dually encoded semi-supervised anomaly detection(DE-SAD)model is proposed. DE-SAD can make full use of a small amount of available labeled data and a large amount of unlabeled data for semi-supervised learning,and learn more accurate implicit manifold distribution of normal data through the dually encoded stage constraint,thus effectively magnifying the gap between normal data and abnormal data. DE-SAD shows excellent ano-maly detection performance on multiple anomaly detection datasets from different fields,especially on low-dimensional data,and its AUROC is up to 4.6% higher than the current state-of-the-art methods.

Keywords Anomaly detection,Semi-supervised learning,Autoencoder,Low-dimensional data

1 引言

异常检测是机器学习领域的一个重要研究问题,旨在检测与大多数数据实例显著偏离的数据实例[1]。异常检测技术被广泛应用于入侵检测、欺诈检测[2-3]、医疗异常检测[4-5]、恶意软件检测[6]、工业异常检测[7-8],以及视频异常检测[9]等领域。

根据训练数据是否有标签,现有异常检测方法大致可分为3类:有监督异常检测、无监督异常检测、半监督异常检测。有监督异常检测使用有标记数据训练有监督分类预测模型,如K近邻(K-nearest neighbors,KNN)、随机森林(Random Forest)和支持向量机(Support Vector Machine,SVM)等。然而,在实际应用中,由于缺乏足够多的可用有标记训练数据,这类方法的可用性较差。无监督异常检测根据数据的内在属性对异常值进行检测,无需标签信息。传统无监督异常检测方法发展比较成熟,如单类支持向量机(One-Class Support Vector Machine,OCSVM)[10-11]、支持向量数据描述(Support

到稿日期:2022-09-05 返修日期:2022-11-30
基金项目:上海市浦江人才计划项目(20PJ1414300);国家自然科学基金联合基金重点支持项目(U1936205);科技部重点研发计划(2021YFC3300300)
This work was supported by the Shanghai Pujiang Program(20PJ1414300),National Natural Science Foundation of China(U1936205) and National Key R&D Program of China(2021YFC3300300).
通信作者:李文根(lwengen@tongji.edu.cn)

Vector Data Description, SVDD)[12]、鲁棒主成分分析(Robust Principle Component Analysis, rPCA)[13]、隔离森林(Isolation Forest, IForest)[14]、局部异常因子算法(Local Outlier Factor, LOF)[15]等。然而,由于无法捕捉数据中的复杂非线性结构,这类方法难以扩展到大规模数据集。近年来,学术界相继提出了许多深度无监督异常检测模型,如自编码器(Auto-Encoder, AE)、变分自编码器(Variational Auto-Encoder, VAE)[16-17]、鲁棒自编码器(Robust Auto-Encoder)[18]、深度自编码高斯混合模型(Deep Autoencoding Gaussian Mixture Model, DAGMM)[19]、深度支持向量数据描述(Deep Support Vector Data Description, Deep SVDD)[20]等。与传统无监督异常检测模型相比,这些方法可以从数据中自动学习深层判别特征,但大多需要对异常数据分布进行先验假设[21],在处理噪声数据时的鲁棒性较差。此外,无监督异常检测模型通常无法结合数据标签进行训练,难以利用可用的有标记数据进一步提升检测性能。半监督异常检测可弥补上述方法的不足,同时利用有标记数据和无标记数据训练模型。例如,Ruff等[22]提出了一种深度半监督异常检测模型 Deep SAD,该模型同时利用大量无标记数据以及可获得的少量标记异常数据进行训练,有效提升了模型的异常检测性能。但该模型更适用于高维数据集的异常检测,如图像异常检测。在实际生产实践中,由于数据特征的获取需要的成本很高,可获取的特征可能非常有限,现有半监督异常检测方法通常难以基于有限的特征挖掘足够的信息用于学习准确数据分布边界。因此,如何利用低维数据进行异常检测是目前亟需解决的问题。

针对上述问题,提出了双编码半监督异常检测模型(DE-SAD),该模型利用少量有标记数据和大量无标记数据,以半监督学习训练深度异常检测神经网络模型。模型采用编码器-解码器-编码器结构,通过半监督学习和双编码约束,能有效减小正常数据的重构误差,增大异常数据的重构误差,从而提升检测性能。由于模型基于数据重构进行异常检测,在低维数据上相比学习数据分布边界的异常检测方法更具优势,不需要大量数据特征来保证边界的准确性。经过实验验证,DE-SAD 模型的异常检测性能在低维数据集上均优于现有方法,在高维数据集上也能表现出较好的检测性能。

2 相关工作

根据使用数据是否有标签,现有异常检测方法大致可以分为 3 类:有监督异常检测、无监督异常检测和半监督异常检测。随着深度学习的发展,这 3 类异常检测方法也逐渐由传统机器学习方法过渡到了深度学习方法。

2.1 有监督异常检测

有监督异常检测使用包含正常数据和异常数据的数据集训练二分类器或多分类器,学习有标记数据的合理决策边界。典型的检测方法包括 K 近邻(KNN)、随机森林(Random Forest)、极限梯度提升树(XGBoost)、支持向量机(SVM)。当有标记训练数据充足时,有监督异常检测的检测性能优于无监督异常检测和半监督异常检测[23]。然而,在实际应用中,通常无法获取到足够多的有标记数据,尤其是异常标记数据,有监督异常检测无法充分学习数据的真实分布,应用范围较小。

2.2 无监督异常检测

无监督异常检测方法应用比较广泛,经典的模型有 OCSVM[10-11]SVDD[12], rPCA[13], IForest[14], LOF[15],以及基于聚类的无监督异常检测算法等。这些传统无监督异常检测方法的主要缺点在于难以挖掘数据特征的复杂结构和非线性关系,导致其检测性能不佳。近年来,基于深度学习的异常检测发展迅速[21]。由于缺少有标记数据,深度异常检测通常基于无监督学习,包括自编码器[24]、变分自编码器和生成对抗网络[25]。自编码器是一种常用的无监督深度学习模型,由编码器和解码器构成,编码器将输入样本数据压缩得到隐含表示特征,解码器将隐含表示特征解码重构。自编码器通过编码和解码的过程学习从输入到自身的映射,可以学习正常数据的特征表示。在训练好的自编码器中,正常数据能被有效重构,而异常数据不能被有效重构。数据样本的重构误差大于设定阈值时,则将其判定为异常值。尽管自编码器可以取得较优的检测性能,但它在部分异常数据上计算的重构误差与正常数据的重构误差非常接近,导致这部分异常数据容易被误检[26]。变分自编码器使用重构概率替换重构误差计算数据的异常分数[16-17],在特定数据集上可以取得更好的异常检测结果。Zhou 等[18]受到鲁棒主成分分析[13]的启发,提出了一种鲁棒自编码器。训练过程中,鲁棒自编码器将含有异常数据的数据矩阵分为低秩矩阵和稀疏的噪声矩阵,其中噪声矩阵中的非零数据被视为异常数据。这种方法的思路比较新颖,然而每次检测与模型训练同时进行,实用性不强。Zong 等[19]提出了 DAGMM 模型,该模型将数据通过自编码器压缩至低维特征空间,并将低维特征表示和重构误差作为新的特征送入混合高斯模型,以端到端的方式联合优化自编码器和高斯混合模型的网络参数,利用混合高斯模型对数据样本进行预测。Ruff 等[20]提出了 Deep SVDD。SVDD 将输入训练数据通过核函数非线性变换映射至一个超球体空间内,并使该超球体半径尽可能小。Deep SVDD 将 SVDD 在神经网络中建模,用神经网络替代核函数提取特征,在高维数据集上表现出了较好的异常检测性能。上述深度无监督异常检测方法可以通过神经网络提高数据深层非线性特征的提取能力,取得更优的检测性能。然而,与有监督异常检测和半监督异常检测相比,深度无监督异常检测方法的主要缺点是不能充分利用可用的有标记数据中蕴含的先验信息进一步提高检测性能。

2.3 半监督异常检测

Görnitz 等[23]早先提出了一种半监督异常检测模型 SSAD,该模型在 SVDD 基础上引入部分有标记数据进行训练,可以取得更好的检测性能。Ruff 等[22]在提出无监督 Deep SVDD 后,又提出了一种端到端的深度半监督异常检测模型 Deep SAD,同时使用大量无标记数据和少量有标记数据进行训练,在不同数据集上取得了较好的检测效果。然而,这类将正常数据压缩至某一超球体的方法,要求超球体有较高的维度来保证超球体的边界足够准确,更适用于高维数据集。当数据特征维度较低时,此类方法难以学习准确的超球边界,导致检测准确度降低。

3　双编码半监督异常检测模型

3.1　模型结构

图 1 给出了双编码半监督异常检测模型(DE-SAD)的基本结构。DE-SAD 模型采用编码器-解码器-编码器结构，其中两个编码器结构相同，但网络参数不同。模型以半监督学习的形式进行训练，输入样本数据 X 由大量无标记数据 X_U 和少量带标签数据 X_L 组成。给定样本数据 $x\in X$，经编编码器 A 压缩至隐含嵌入层，得到隐含特征 z，z 经解码器解码重构为 $\hat{x}$，$\hat{x}$ 与 x 维数相同。而后 $\hat{x}$ 经编码器 B 再次压缩至另一隐含嵌入层，得到隐含特征 $\hat{z}$，$\hat{z}$ 与 z 维数相同。上述过程的公式描述如下所示：

$$z=\phi_{e_A}(x,\theta_{e_A}) \tag{1}$$

$$\hat{x}=\phi_d(z,\theta_d) \tag{2}$$

$$\hat{z}=\phi_{e_B}(\hat{x},\theta_{e_B}) \tag{3}$$

其中，φ_{e_A}，φ_d 和 φ_{e_B} 分别表示编码器 A、解码器和编码器 B；θ_{e_A}，θ_d 和 θ_{e_B} 为对应的网络参数。

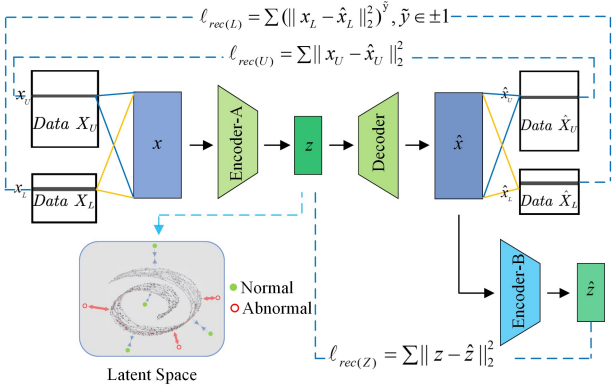

图 1　双编码半监督异常检测模型结构

Fig. 1　Structure of dually encoded semi-supervised anomaly detection model

3.2　半监督学习

对于无标记数据 X_U，模型通过最小化数据的重构误差学习输入数据到其自身的映射，使编码器 A 可以获得正常数据隐含的特征分布，表示如下：

$$\min_{\theta_{e_A},\theta_d}\sum\|x_U-\hat{x}_U\|_2^2 \tag{4}$$

其中，$\hat{x}_U$ 为 x_U 的重构样本，θ_{e_A} 和 θ_d 为编码器 A 和解码器的网络参数。

为有效利用可获得的少部分有标记数据来提升模型的异常检测性能，DE-SAD 模型结合这部分有标记数据和大部分无标记数据进行半监督学习训练，通过有标记数据的先验信息使模型更准确地获得正常数据的潜在分布。对于有标记数据，优化目标为：

$$\min_{\theta_{e_A},\theta_d}\sum(\|x_L-\hat{x}_L\|_2^2)^{\tilde{y}},\tilde{y}\in\pm1 \tag{5}$$

其中，$\hat{x}_L$ 为 x_L 的重构样本；θ_{e_A} 和 θ_d 为编码器 A 和解码器的网络参数；$\tilde{y}=+1$ 表示正常数据，$\tilde{y}=-1$ 表示异常数据。当有标记数据为正常数据时，模型对其重构误差施加二次项损失；当有标记数据为异常数据时，模型对其重构误差的倒数进行惩罚，从而引导模型在低维流形层面进一步将正常数据拉近流形表面，将异常数据推离流形表面。

3.3　双编码约束

为进一步约束模型学习更准确稳定的潜在分布表示，模型将两个编码器输出的隐含嵌入层引入反向传播优化过程中。由于隐含嵌入层可以获得正常数据的低维潜在分布，因此随着训练的收敛，两个隐含嵌入层的嵌入表示越来越接近正常数据的真实分布，二者逐渐趋近。优化目标表示为：

$$\min_{\theta_{e_A},\theta_d,\theta_{e_B}}\sum\|z-\hat{z}\|_2^2 \tag{6}$$

其中，z 和 $\hat{z}$ 分别表示编码器 A 和编码器 B 输出的隐含嵌入层表示；θ_{e_A}，θ_d 和 θ_{e_B} 为解码器 A、编码器和解码器 B 的参数。

3.4　损失函数

综合考虑式(4)－式(6)，DE-SAD 模型最终的目标损失函数为：

$$\frac{1}{M}\sum_{i=1}^{M}\|x_i-\hat{x}_i\|_2^2+\frac{\eta_1}{N}\sum_{j=1}^{N}(\|x_j-\hat{x}_j\|_2^2)^{\tilde{y}_j}+\frac{\eta_2}{M+N}\sum_{k=1}^{M+N}\|z-\hat{z}\|_2^2+\frac{\lambda}{2}\sum_{l=1}^{L}\|W^l\|_F^2 \tag{7}$$

其中，M 表示无标记数据的数量；N 表示有标签数据的数量；$\tilde{y}_j\in\pm1$ 为有标签数据的标签；W 为模型全局隐含层网络权重参数；L 是模型隐含层层数；$\|\cdot\|_F$ 表示 Frobenius 矩阵范数，在综合考虑式(4)－式(6)的基础上，通过 Frobenius 矩阵范数对全局网络权重进行正则化约束，实现网络权重衰减，防止模型训练过拟合；η_1，η_2 和 λ 为超参数，用来调节各部分损失函数及正则项的权重。

3.5　模型训练与测试

DE-SAD 模型训练阶段算法实现过程如算法 1 所示，其中 epochs 表示训练轮次，batch_size 表示批量大小。首先，对网络参数随机初始化，然后在每一个训练轮次中，遍历每一个批次，按批次将每个数据样本输入模型网络中，通过小批量梯度下降和反向传播优化模型网络。

算法 1　DE-SAD 模型训练

输入：训练集 $X=\{x_i\}_{i=1}^{M+N}$，$(M>N)$

输出：编码器 A(φ_{e_A})、解码器(φ_d)和编码器 B(φ_{e_B})

1. 随机初始化 φ_{e_A}，φ_d 和 φ_{e_B} 的网络参数 θ_{e_A}，θ_d 和 θ_{e_B}
2. 设定 epochs 和 batch_size 大小
3. iterations←⌈(M+N)/batch_size⌉
4. for epoch←1 to epochs do：
5.　　for i←1 to iterations do：
6.　　for j←1 to batch_size do：
7.　　通过式(1)计算获得训练样本 x_j 经过编码器 A 编码后得到的隐含嵌入特征 z_j
8.　　通过式(2)计算获得 z_j 经解码器解码重构输出的 $\hat{x}_j$
9.　　通过式(3)计算获得 $\hat{x}_j$ 经编码器 B 编码后得到的隐含嵌入特征 $\hat{z}_j$
10.　　通过式(7)反向传播求解梯度，更新网络参数 θ_{e_A}，θ_d 和 θ_{e_B}
11.　end for
12.　end for
13. end for

在测试阶段，输入测试数据的异常分数为输入测试数据 x_t 重构误差的平方，即

$$s(x_t)=\| x_t-\hat{x}_t \|_2^2 \tag{8}$$

其中，$\hat{x}_t$ 为解码器对 x_t 隐含嵌入特征的重构输出。训练好的模型将测试数据编码至正常数据的隐含层特征表示，而后正常数据和异常数据均会被解码器映射为正常数据。因此，正常数据的异常分数非常小，异常数据会获得较大的异常分数。与自编码器相比，DE-SAD 模型可以使正常数据的异常分数更小，并放大异常数据的异常分数，从而提升对异常数据的检出性能。

测试阶段的算法实现相对比较简单，如算法 2 所示。测试阶段只需使用编码器 A 和解码器计算测试样本的异常分数，当异常分数高于设定阈值时，该测试样本即被判定为异常。阈值用于调节被检出异常样本的多少，阈值越小，被检出的异常样本就越多，但假正类率会随之升高。通过调整阈值可以得到不同的异常检测结果。

算法 2 基于 DE-SAD 模型的异常检测

输入：测试数据 x_t，已训练好的编码器 $A(\varphi_{e_A})$、解码器(φ_d)和阈值 δ

输出：x_t 的异常分数及其类别

1. 利用式(1)计算得到 x_t 经编码器 A 编码后输出的隐含嵌入特征 z_t
2. 利用式(2)计算得到 z_t 经解码器解码重构输出的 $\hat{x}_t$
3. 利用式(8)计算 x_t 的异常分数 $s(x_t)$
4. if $s(x_t)<\delta$:
5. then:
6. x_t 是正常数据
7. else:
8. x_t 是异常数据
9. end if

4 实验与分析

4.1 数据集及其预处理

为检验所提模型的异常检测性能，实验选用了多组数据集：食品污染物检测数据集、可饮用水污染物检测数据集[1)]，以及来自其他不同领域的多个标准异常检测数据集[27]。所有数据集均为结构化表格数据。食品污染物检测数据集包含 8 个理化指标，该数据集包含正常食品样本和异常样本两类。可饮用水污染物检测数据集来自 Kaggle 平台，包含 9 个测量的理化特征，分为可饮用水和不可饮用水两类。所有数据集的详细信息如表 1 所列，包括总样本数、异常样本数、特征维度，以及异常样本占比。这些数据集的特征维度大多较少，且异常样本比例都较低。

表 1 实验数据集详细信息

Table 1 Dataset details for experiments

dataset	# samples	# anomalies	# features	# abnormal proportion/%
food	66856	466	8	0.7
water	1220	20	9	1.6
pima	510	10	8	2.0
annthyroid	6766	100	6	1.5
thyroid	3772	93	6	2.5
shuttle	49097	3511	9	7.2
satimage	5803	71	36	1.2
mnist-100	7003	100	100	1.4

1) https://www.kaggle.com/code/creepycrap/potable-water-71-acc/data

实验数据集均统一进行预处理操作。首先将样本标签转化为±1，其中+1 表示正常样本，−1 表示异常样本。然后，对特征数据进行标准化处理，使其符合标准正态分布，转化公式为：

$$X^*=\frac{X-\mu}{\sigma} \tag{9}$$

其中，X 表示原始数据样本，μ 和 σ 分别表示数据的算术平均值和标准差，X^* 为标准化处理后的数据样本。

最后，将经过标准化处理后的数据进行特征缩放，归一化至[−1,1]区间，有利于加快模型梯度下降的速度。特征缩放公式为：

$$\tilde{X}=\frac{X^*}{\max\{|X^*|\}} \tag{10}$$

实验中随机抽取 20%的数据作为测试集，剩余 80%的数据作为训练集。

4.2 对比方法与评价指标

实验选取隔离森林(IForest)、单类支持向量机(OCSVM)、鲁棒主成分分析(rPCA)[13]、自编码器(AE)、深度支持向量数据描述(Deep SVDD)，以及半监督异常检测模型 Deep SAD 等主流异常检测模型进行对比。实验评价指标为 ROC 曲线下方面积(Area Under ROC, AUROC)。AUROC 是综合评价不同阈值下各模型异常检测性能的通用指标。为避免随机性的影响，每种对比方法在每个数据集上均取 10 次不同的随机种子，对于某个固定的随机种子，重复进行 5 次实验，每种实验设置下一共进行 50 组实验，得到 50 个 AUROC 结果并计算平均值。

4.3 模型配置

实验环境设置如下：使用深度学习框架 PyTorch 进行编程，实验在 Linux 服务器上进行，使用 NVIDIA 2080Ti 加速。实验中使用 Adam 优化器[28]并设置激活函数为 Tanh 函数对 DE-SAD 模型进行训练。对于不同数据集，DE-SAD 模型的编码器结构根据数据集特征维度分别进行设置，解码器与编码器对称。通过实验验证分析，在不同数据集上网络及超参数的最佳设置如表 2 所列。每组实验均训练 100 个 epoch，在训练过程中，每 50 个 epoch 后学习率以 10 倍率递减。如此，当模型损失收敛后学习率自动微调，可以得到更好的训练结果。对比模型的超参数设置则参考对应文献，并在不同数据集上进行微调。

表 2 DE-SAD 模型参数设置

Table 2 Parameter setting of DE-SAD

dataset	encoder parameters	learning rate	batch size	η_1	η_2	λ
food	{8,5,3}	0.0020	1024	1	1	0.5×10^{-5}
water	{9,6,3}	0.0020	16	1	1	0.5×10^{-5}
pima	{8,5,3}	0.0010	16	1	1	0.5×10^{-5}
annthyroid	{6,4,3}	0.0020	64	1	1	0.5×10^{-6}
thyroid	{6,4,3}	0.0020	64	1	1	0.5×10^{-6}
shuttle	{9,6,3}	0.0020	1024	1	1	0.5×10^{-5}
satimage	{36,24,10}	0.0020	64	1000	1	0.5×10^{-6}
mnist-100	{100,50,24}	0.0005	256	1000	1	0.5×10^{-5}

4.4　结果与分析

4.4.1　不同异常检测模型对比

各模型在不同数据集上的检测结果如表 3 所列，表中数据为 50 组实验的 AUROC 平均值，附以 95% 置信区间的抖动值。表 3 中最优结果用粗体标出，次优结果用下划线标出。由结果可以看出，DE-SAD 模型在大部分数据集上都表现出最优的异常检测性能，相比次优方法，DE-SAD 的 AUROC 平均值最高提升了 4.6%，并且其结果抖动值区间较小，表明所提模型鲁棒性较好。当数据集的特征维度较高时，如在 mnist-100 数据集上，DE-SAD 表现出次优的性能，但相比传统方法及自编码器和 Deep SVDD 具有明显优势。Deep SAD 通过学习正常数据的超球体边界来进行异常检测，因此其在较高维数据上学习到的正常数据边界更接近真实边界。DE-SAD 基于重构进行异常检测，在低维数据集上更具有优势。就模型复杂度而言，Deep SAD 的训练过程分为预训练和训练两阶段，其通过自编码器预训练获得模型初始化参数后再开始训练模型，而 DE-SAD 不用预训练，模型复杂度大大降低。综合上述分析，和现有主流异常检测模型相比，DE-SAD 有明显的异常检测优势，尤其是在低维数据上。

表 3　不同异常检测模型在多个数据集上的测试结果

Table 3　Test results of different anomaly detection methods on multiple datasets

（单位：%）

Dataset	IForerst	OCSVM	rPCA	AE	Deep SVDD	Deep SAD	DE-SAD
food	72.3±0.8	70.8±0.9	74.3±0.7	73.2±0.7	63.6±0.4	75.5±0.8	**78.5±0.7**
water	73.4±2.0	69.7±2.3	72.7±2.0	71.1±2.0	61.4±5.1	72.0±3.0	**76.6±1.9**
pima	79.8±2.4	77.6±4.8	80.3±2.6	79.9±3.9	66.8±6.9	79.5±4.8	**83.5±2.5**
annthyroid	90.2±1.0	66.8±2.0	80.6±1.8	58.5±4.2	54.7±3.1	77.2±3.9	**93.4±1.2**
thyroid	97.7±0.2	92.6±1.1	96.3±0.4	86.0±1.9	53.1±4.5	84.2±7.2	**98.5±0.1**
shuttle	99.1±0.1	90.0±3.7	98.3±0.1	95.2±1.6	87.3±4.9	99.0±0.2	**99.2±0.1**
satimage	98.8±0.3	88.6±1.1	96.7±0.8	96.4±0.9	89.3±3.2	99.5±0.2	**99.6±0.2**
mnist-100	83.5±1.2	83.4±1.5	87.9±1.1	90.5±1.0	71.3±2.2	**94.1±0.9**	93.1±0.9

4.4.2　消融分析

为验证 DE-SAD 中半监督学习机制和双编码约束的有效性，本文以食品污染物检测数据集为例进行了消融实验。图 2 给出了 DE-SAD 消融实验的结果，每个箱线图对 50 组实验结果进行统计描述。其中 DE-UAD 表示双编码无监督异常检测模型（Dually Encoded Unsupervised Anomaly Detection），即完全使用无标记数据进行无监督训练；SE-SAD 表示单编码半监督异常检测模型（Singly Encoded Semi-supervised Anomaly Detection），相比 DE-SAD，其网络主干将编码器 B 移除。

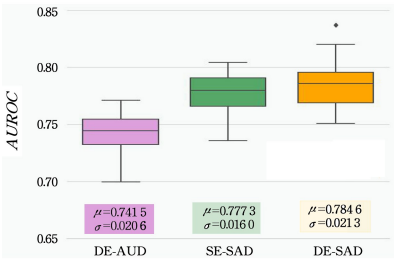

图 2　消融实验结果

Fig. 2　Results of ablation experiments

由图 2 箱线图可观察到，SE-SAD 的 AUROC 平均值为 77.73%，相较于 DE-SAD 有所下降，验证了模型双编码约束阶段的有效性。DE-UAD 的 AUROC 平均值为 74.15%，相较于 DE-SAD 下降明显，验证了模型半监督学习机制的有效性。

4.4.3　有标记数据占比调节

为验证训练集中有标记数据占比对两种半监督模型检测性能的影响，在食品污染物数据集上使用 Deep SAD 模型和 DE-SAD 模型进行了不同比例有标记数据下的实验分析。实验结果如图 3 所示，其中不同颜色表示训练集中有标记数据的不同占比。随着有标记数据占比逐渐增大，Deep SAD 和 DE-SAD 的 AUROC 平均值逐渐升高。实验结果表明，在训练集中加入有标记数据即可有效提高 DE-SAD 模型的检测性能，且在相同实验条件下，DE-SAD 的检测性能显著优于 Deep SAD。

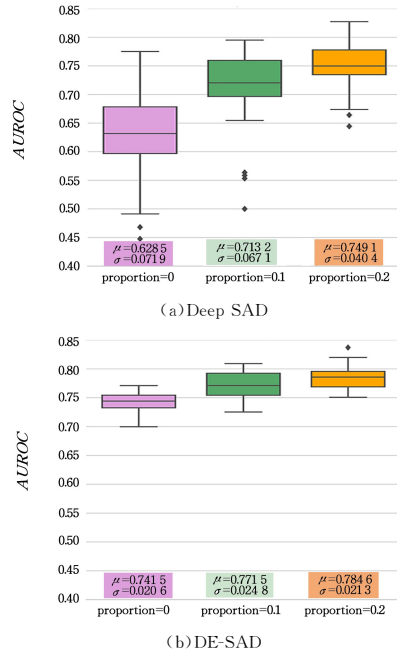

图 3　有标记数据占比调节的结果（电子版为彩图）

Fig. 3　Results of adjusting the proportion of labeled data

4.4.4　异常分数分布演化

为直观呈现 DE-SAD 模型的检测性能，以食品污染物数据集为例，对训练集在训练过程中不同阶段的异常分数分布进行绘图。训练集异常分数分布演化情况如图 4 所示，其中绿色表示正常数据，蓝色表示异常数据，横轴表示异常分数。初始阶段，正常数据和异常数据分布基本重叠，难以有效

区分。随着训练过程迭进，正常数据和异常数据的异常分数分布出现差距，且差距逐渐明显，最终异常数据的异常分数高于正常数据的异常分数。实验结果对训练过程中数据异常分数分布的演化进行了直观描述，并说明了模型能有效拉大正常数据和异常数据的差距。

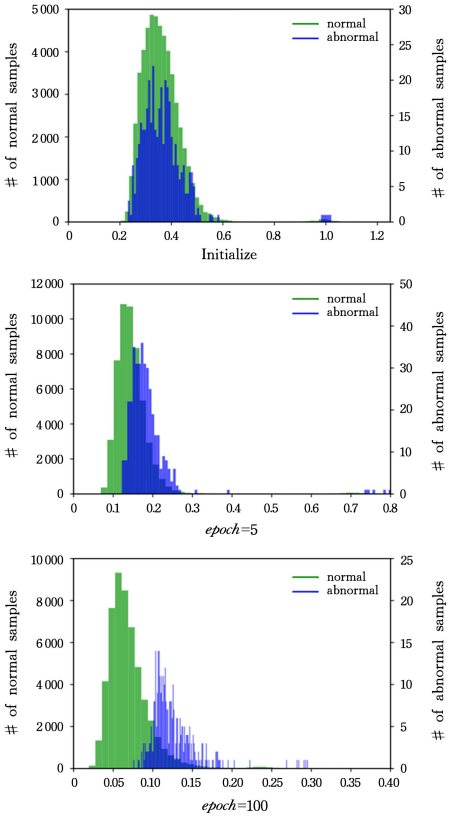

图 4 异常分数分布演化(电子版为彩图)

Fig. 4 Evolution of anomaly score distribution

结束语 针对目前主流的异常检测算法无法充分利用可获得的少量有标记数据来提升异常检测性能，且在低维数据上检测效果不足的问题，本文提出了双编码半监督异常检测模型(DE-SAD)。DE-SAD 基于半监督学习，利用少部分有标记数据结合大量无标记数据进行联合训练，通过双编码阶段约束模型学习到更准确的正常数据隐含分布流形表示。模型基于数据重构进行异常检测，能在最小化正常数据的重构误差的同时放大异常数据的重构误差。DE-SAD 在来自不同领域的多个异常检测数据集上都表现出优异的异常检测性能。消融实验验证了 DE-SAD 的半监督学习机制以及双编码约束阶段的有效性，有标记数据占比调节实验验证了有标记数据对模型的检测性能具有显著提升作用。可视化实验则展示了训练过程中数据异常分数的分布变化情况，对模型的训练过程进行了形象展示。在未来的研究中，可以着重针对高维数据进行模型优化，使模型能适用于更多异常检测数据集，进一步提高模型的通用性。

参 考 文 献

[1] PANG G, SHEN C, CAO L, et al. Deep Learning for Anomaly Detection: A Review [J]. ACM Computing Surveys, 2021, 54(2): 38: 1-38: 38.

[2] ILEBERI E, SUN Y, WANG Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection [J]. Journal of Big Data, 2022, 9(1): 1-17.

[3] BIN S R, SCHETININ V, SANT P. Review of Machine Learning Approach on Credit Card Fraud Detection[J]. Human-Centric Intelligent Systems, 2022, 2(1/2): 55-68.

[4] LI M M, HUANG K, ZITNIK M. Graph representation learning in biomedicine and healthcare[J]. arXiv: 2104. 04883, 2022.

[5] WANG J, JIA Y, WANG D, et al. Weighted IForest and siamese GRU on small sample anomaly detection in healthcare[J]. Computer Methods and Programs in Biomedicine, 2022, 218: 106706.

[6] CHAGANTI R, RAVI V, PHAM T D. Deep learning based cross architecture internet of things malware detection and classification[J]. Computers & Security, 2022, 120: 102779.

[7] DE PAULA MONTEIRO R, LOZADA M C, MENDIETA D R C, et al. A hybrid prototype selection-based deep learning approach for anomaly detection in industrial machines[J]. Expert Systems with Applications, 2022, 204: 117528.

[8] KHARITONOV A, NAHHAS A, POHL M, et al. Comparative analysis of machine learning models for anomaly detection in manufacturing[J]. Procedia Computer Science, 2022, 200: 1288-1297.

[9] ZAHEER M Z, MAHMOOD A, KHAN M H, et al. Generative cooperative learning for unsupervised video anomaly detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14744-14754.

[10] SCHÖLKOPF B, PLATT J C, SHAWE-TAYLOR J, et al. Estimating the Support of a High-Dimensional Distribution [J]. Neural Computation, 2001, 13(7): 1443-1471.

[11] WU Y K, LI W, NI M Y, et al. Anomaly Detection Model Based on One-Class Support Vector Machine Fused Deep Autoencoder [J]. Computer Science, 2022, 49(3): 144-151.

[12] TAX D M, DUIN R P. Support vector data description[J]. Machine Learning, 2004, 54(1): 45-66.

[13] SHYU M L, CHEN S C, SARINNAPAKORN K, et al. A novel anomaly detection scheme based on principal component classifier[C]// Proceedings of the 3rd IEEE International Conference on Data Mining. 2003: 172-179.

[14] LIU F T, TING K M, ZHOU Z H. Isolation Forest[C]// 2008 Eighth IEEE International Conference on Data Mining. 2008: 413-422.

[15] CHENG Z, ZOU C, DONG J. Outlier detection using isolation forest and local outlier factor[C]// Proceedings of the Conference on Research in Adaptive and Convergent Systems. 2019: 161-168.

[16] ZHANG R J, CHEN W, HANG M X, et al. Detection of Abnormal Flow of Imbalanced Samples Based on Variational Autoencoder[J]. Computer Science, 2021, 48(7): 62-69.

[17] CHEN Q, DAI Y, LIU G. Research on KPI Anomaly Detection Model for Intelligent Operation and Maintenance[J]. Journal of Chongqing University of Technology (Natural Science), 2022, 36(6): 181-188.

[18] ZHOU C, PAFFENROTH R C. Anomaly Detection with Robust Deep Autoencoders[C]// Proceedings of the 23rd ACM

SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax NS Canada: ACM, 2017: 665-674.

[19] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]// International Conference on Learning Representations. 2018: 1-19.

[20] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification[C]// International Conference on Machine Learning. PMLR, 2018: 4393-4402.

[21] CHALAPATHY R, CHAWLA S. Deep Learning for Anomaly Detection: A Survey[J]. arXiv: 1901. 03407, 2019.

[22] RUFF L, VANDERMEULEN R A, GÖRNITZ N, et al. Deep Semi-Supervised Anomaly Detection[J]. arXiv: 1906. 02694, 2020.

[23] GÖRNITZ N, KLOFT M, RIECK K, et al. Toward supervised anomaly detection[J]. Journal of Artificial Intelligence Research, 2013, 46: 235-262.

[24] YUAN F N, ZHANG L, SHI J T, et al. Review of Autoencoder Neural Network Theory and Applications [J]. Journal of Computers, 2019, 42(1): 203-230.

[25] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. Ganomaly: Semi-supervised anomaly detection via adversarial training[C]// Asian Conference on Computer Vision. Springer, 2018: 622-637.

[26] GONG D, TAN M, ZHANG Y, et al. Blind Image Deconvolution by Automatic Gradient Activation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1827-1836.

[27] ODDS-Outlier Detection DataSets [EB/OL]. http://odds.cs.stonybrook.edu/.

[28] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv: 1412. 6980, 2014.

LI Hui, born in 1997, postgraduate, is a member of China Computer Federation. His main research interests include anomaly detection and big data analysis.

LI Wengen, born in 1987, Ph. D, assistant professor, is a member of China Computer Federation. His main research interest is spatio-temporal data management and analysis.

（责任编辑：何杨）