中国科学: 信息科学 2022 年 第 52 卷 第 1 期: 75–85
SCIENTIA SINICA Informationis
论文

连续无监督异常检测

倪一鸣[1,2], 陈松灿[1,2*]

1. 南京航空航天大学计算机科学与技术学院, 南京 210016
2. 模式分析与机器智能工业和信息化部重点实验室, 南京 210016
* 通信作者. E-mail: s.chen@nuaa.edu.cn

收稿日期: 2021–05–31; 修回日期: 2021–08–25; 接受日期: 2021–09–18; 网络出版日期: 2022–01–10

航空发动机及燃气轮机重大专项基础研究项目 (批准号: J2019-IV-0018-0086) 和国家自然科学基金 (批准号: 62076124) 资助项目

摘要 无监督异常检测 (unsupervised anomaly detection, UAD) 旨在检测任何未见过的偏离预期模式或正常分布的数据, 由于其学习过程不依赖对罕见异常样本的获取, 因此在现实动态环境下备受青睐. 然而, 在现实场景中, 目标任务往往会随时间动态变化, 这要求模型能够连续执行多个不同的 UAD 任务, 确保在仅有当前任务正常数据的前提下, 实现对所有见过任务的异常检测. 本文旨在研究这一问题, 尝试从互信息角度, 提出一种新的连续 UAD (CUAD) 算法. 具体而言, 我们针对原始目标依赖过往任务原始数据和异常数据的问题, 给出基于信息论的损失函数, 并对其进行近似优化. 据此, 我们构建出来的深度编码器模型既能连续执行不同的 UAD 任务, 又能有效应对连续学习带来的灾难性遗忘问题. 最后, 我们在多个标准数据集上的实验验证了所提出方法的优越性.

关键词 异常检测, 无监督, 灾难性遗忘, 连续学习, 信息论

1 引言

异常检测 (anomaly detection, AD) 旨在检测与预期模式或正常分布不符的异常数据 [1], 在诸如医学图像诊断 [2]、信用卡金融诈骗 [3]、网络安全 [4]、传感器异常 [5] 等多个领域得到应用, 因此已产生了丰富的理论与应用成果 [6]. 在现实场景下, 正常数据通常大量存在且易于获得, 而异常数据则相对不易获得, 且常常缺失, 同时还往往呈现出多样性 [7], 因此, 相对于需要采用标记正常和异常数据的监督问题, 仅利用正常数据的异常检测自然成为了首选的策略. 此类策略通常被称为无监督异常检测 (unsupervised anomaly detection, UAD) 或单分类学习问题 [8] (one-class classification, OCC). 目前已有众多学习模型及相应算法被提出 [9~11], 尤其近年来, 深度神经网络 (deep neural network, DNN) 凭借其极强的学习和特征表征能力, 在计算机视觉、自然语言处理等领域取得很好的效果 [12], 并在物体识别 [13] 和围棋 [14] 等任务上超越了人类. 伴随着深度学习的蓬勃发展, 各类深度 UAD 也相继被提出

引用格式: 倪一鸣, 陈松灿. 连续无监督异常检测. 中国科学: 信息科学, 2022, 52: 75–85, doi: 10.1360/SSI-2021-0192
Ni Y M, Chen S C. Continual unsupervised anomaly detection (in Chinese). Sci Sin Inform, 2022, 52: 75–85, doi: 10.1360/SSI-2021-0192

且性能优异 [15,16], 这些方法主要可分为基于自编码器的模型和混合模型. 前一类模型主要依据重建误差以判定异常 [17], 而混合模型往往将自编码器与传统异常检测模型级联 (如 OC-SVM [18,19]) 达成有效检测.

进一步, UAD 根据应用场景的连续与否, 又可分为静态与动态检测. 静态 UAD 指单时刻 UAD 任务 [20]. 而动态 UAD 则是针对连续的多个检测任务, 即连续执行一系列 UAD 任务, 各任务拥有自身的正常数据与未知的异常数据, 当前任务的模型仅学习其正常类, 无需重训练过往任务, 便可有效地识别和检测所有见过的正常和异常模式 [21]. 这种针对动态场景的 UAD 称为连续无监督异常检测 (continual unsupervised anomaly detection, CUAD), 即本文关注的主题.

在现实中存在众多此类典型场景, 如工厂中随着配置的传感器数的增加与更新, 要求 AD 算法无需在过往数据上进行繁琐的重训练, 即可快速实现对新老传感器异常的有效检测; 自动驾驶任务中, 面对瞬息多变的真实环境, 甚至不可预知的异常, 力争有效而准确地作出检测.

CUAD 问题的独特之处在于异常样本的缺失与灾难性遗忘 (catastrophic forgetting) [22~24], 并由此带来两大挑战: (1) 寻找更为有效和适应无监督学习的损失函数, 以替代因异常类缺少而失效的常用典型损失; (2) 克服缺少过往数据以及模型在新任务上过拟合导致的灾难性遗忘.

针对 CUAD 的两大挑战, 最近, Frikha 等 [21] 提出了借用元学习 [25] 思想, 通过在大量异常检测任务上的训练, 得到较好的初始化参数和学习率, 并将其作为元知识以缓解灾难性遗忘. 该方法尽管获得了较好的效果, 但未能考虑执行连续任务时参数的动态更新. Wiewel 等 [26] 针对灾难性遗忘这一挑战, 构建了一个自编码器以生成重放 (generative replay) 以往任务的带标签数据, 并将其联合新任务数据共同训练. 但此方法不仅需要使用带标签数据而且设置了较为复杂的生成重放步骤, 增加了模型的标注与计算代价.

现有算法损失函数由于依赖异常数据和过往任务数据, 难以适用于连续无监督场景, 而从信息论的角度设计损失函数, 并分解近似, 可得到不再依赖异常数据 [27] 与原始数据的损失. 因此, 为了克服现有算法的不足, 我们基于 AD 中 “最大化正常数据和异常数据在隐空间中的特征分布间距离” 的思想和信息论来设计关键的损失函数, 进而对其分解近似实现优化, 提出了 CUAD 算法.

本文贡献在于: (1) 从信息论角度设计了一个全新的不再依赖异常数据和原始数据的损失函数; (2) 提出了一个新的连续无监督异常检测算法 CUAD; (3) 在 4 个常用标准数据集上验证了 CUAD 算法相对现有算法表现更优异.

接下来的第 2 节中, 介绍连续学习和异常检测及代表方法; 第 3 节具体介绍适应无监督学习和动态场景的损失函数及 CUAD 算法; 第 4 节中, 在包括 MNIST, CIFAR-FS, MiniImageNet, KDDCup99 的多个常用标准数据集上进行 CUAD 算法的有效性检验, 并与现有算法进行比较, 验证了其优越性; 最后总结全文并展望未来工作.

2 相关工作

2.1 UAD 与单分类问题

为了寻找数据中的异常, AD 算法可分为以下三类: (1) 构建自编码器等模型 [28,29] 重建样本, 正常样本能通过有限的基函数实现重建, 而异常数据会引起较大重建误差, 从而得以检出; (2) 基于样本分布 [30,31], 构建高斯 (Gauss) 混合、核密度估计等概率模型以逼近正常数据的分布, 进而通过最大似然检出异常; (3) 基于决策树 [32,33], 使用孤立树以分割样本, 借助样本在森林中的平均高度判断异常.

对于仅使用正常类数据的 UAD, 可借助流行的分类技术, 将检测问题转化为单分类问题, 如 OC-SVM [19] 等模型. 同时借力深度网络的突出优势, 众多深度模型 AD 也应运而生, 并获得了更佳的性能. 典型的模型有: 深度自编码器 [31] 将原始数据映射到更易检测的特征空间, 生成对抗网络通过重建函数实现 AD [34], 基于迁移学习的预训练网络方法 [35,36], 深度编码的高斯混合模型方法 [37], 基于数据增广的自监督方法 [10,27,38] 等. 这些方法均可缓解无监督条件下的异常数据缺失问题, 但由于考虑的是静态问题, 并不适应连续场景.

2.2 连续学习

连续学习又称终生学习 (lifelong learning) 或增量学习 (incremental learning), 旨在从数据序列中连续学习知识, 并完成一系列任务. 学习新任务会导致模型对过往任务性能上的灾难性退化或遗忘 [22], 针对此问题, 研究者提出了以下多种连续学习策略: (1) 基于重放技术, 如 iCaRL [39], GEM [40,41] 等模型通过存储过往任务中代表性样本, 用于学习新任务时的联合训练, 实现对过往知识的温习; (2) 基于正则化技术, 若无法存储原始数据, 新策略是学习新任务同时强化过往知识, 如 EWC (elastic weight consolidation) [24], LwF [42], SI [43] 等模型在损失函数中添加对关键参数的正则化约束, 抑制模型参数的变化以缓解对过往知识的遗忘; (3) 参数分离技术, 如 PackNet [44], PNN [45] 等模型对每个任务设计专属参数与模型, 从而避免参数更新引起性能的下降, 但此类策略导致模型规模随着任务增加而趋于庞大, 故其应用场景受限; (4) 博取众家之长, 结合上述各类策略 [46~48], 融合多类技术, 以达到更少的对知识的遗忘. 此外, 当任务按类渐增方式学习时, 相应的连续学习被称为类增量学习 (class incremental learning) [39], 本文主要研究连续地执行 UAD 任务, 即连续无监督异常检测.

3 提出模型

3.1 问题定义

设共有 T 个异常检测任务, x_i 代表第 i 个任务的原始数据, $z_i^t = E_{\theta_t}(x_i)$ 表示其在 t 时刻的模型 E_{θ_t} 中所学得的隐特征, 记 $p_n(z_i^t|x_i)$ 和 $p_a(z_i^t|x_i)$ 分别代表正常数据和异常数据隐空间分布在原始空间分布下的条件分布.

当学习第 t 个任务时, 训练数据仅包括当前的正常数据 x_t, 检测任务包括任务 $1 \sim t$ 中所有正常及异常数据.

检测多个任务中的异常数据, 不仅需要考虑当前任务 t 的损失函数 $l(x_t, z_t^t)$, 还包含过往任务 i 中的数据, 针对原始损失依赖过往任务原始数据和异常数据这一问题, 我们考虑从信息论的角度给出损失, 同时应对两类数据缺失的困难, 通过分解近似, 给出新的不依赖这两类数据的损失函数. 故考虑最大化第 i 个任务中正常与异常数据在当前编码器 E_{θ_t} 中隐特征的分布差异, 即最小化

$$L(x, z, \theta) = l(x_t, z_t^t) - \lambda KL[p_n(z_i^t|x_i)||p_a(z_i^t|x_i)], \tag{1}$$

其中, λ 表示正则化参数.

3.2 损失函数

式 (1) 中, $z_i^t = E_{\theta_t}(x_i)$ 表示第 i 个任务数据通过 t 时刻任务中编码器得到的隐特征. 依据无监督学习与连续学习设定, 学习任务 t 时, 仅可获取 x_t 中的正常数据, 无异常数据和过往任务数据. 因此,

对式 (1) 中第 2 项损失进行分解:

$$\begin{aligned}
&\max KL[p_n(z_i^t|x_i)||p_a(z_i^t|x_i)]\\
&=\max \mathbb{E}_{p_n(z_i^t|x_i)}\left[\log\frac{p_n(z_i^t|x_i)}{p_a(z_i^t|x_i)}\right]\\
&=\max \mathbb{E}_{p_n(z_i^t|x_i)}\left[\log\frac{p_n(z_i^t|x_i)}{p_n(z_i^t)}\frac{p_n(z_i^t)}{p_a(z_i^t|x_i)}\right]\\
&=\max \mathbb{E}_{p_n(z_i^t|x_i)}\left[\log\frac{p_n(z_i^t|x_i)p_n(x_i)}{p_n(z_i^t)p_n(x_i)}\right]+\mathbb{E}_{p_n(z_i^t|x_i)}\left[\log\frac{1}{p_a(z_i^t|x_i)}\right]+\mathbb{E}_{p_n(z_i^t|x_i)}[\log p_n(z_i^t)]\\
&=\max I(x_i;z_i^t)-H(z_i^t)+\mathbb{E}_{p_n(z_i^t|x_i)}[-\log p_a(z_i^t|x_i)]\\
&=\max I(x_i;z_i^t)-H(z_i^t)+\mathbb{E}_{p_n(x_i)}[H(p_n(z_i^t|x_i),p_a(z_i^t|x_i))].
\end{aligned}\tag{2}$$

分解得到的式 (2) 中 3 项分别代表: 第 i 个任务中正常数据与当前编码器下编码之间的互信息, 当前编码器映射的隐空间中特征分布的熵, 第 i 个任务中正常数据与异常数据条件概率之间的交叉熵.

式 (2) 中的损失存在 2 个问题: 互信息需要过往任务原始数据, 交叉熵需要异常数据. 文献 [27] 中命题 2 作出合理假设, 约束交叉熵下界为 0. 文献 [49] 中介绍了一种基于噪声对比预测 (noise-contrastive estimation, NCE) 的损失 InfoNCE 以实现互信息的最大化:

$$I_{\mathrm{NCE}}(x_i;z_i^t)=\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\log\frac{\exp(f(x_{ik},z_{ik}^t))}{\frac{1}{K}\sum_{j=1}^{K}\exp(f(x_{ik},z_{ij}^t))}\right],\tag{3}$$

其中 $f(x_{ik},z_{ik}^t)$ 表示第 i 个任务中, 第 k 个数据及其在第 t 个编码器中学得的隐特征之间的一致性. 值得注意的是, 在学习第 i 个任务时, 训练得到的编码器 E_{θ_i} 是使得 X_i 与 Z_i^i 之间损失最小化的编码器, 为了充分利用训练中编码器获取的知识并弥补缺少过往任务数据带来的缺陷, 对一致性判定函数采用此设定:

$$f(x_{ik},z_{ik}^t)=\mathrm{sim}(z_{ik}^t,E_{\theta_i}(x_{ik}))=\mathrm{sim}(z_{ik}^t,z_{ik}^i),\tag{4}$$

其中, $\mathrm{sim}(\cdot,\cdot)$ 代表相似度函数, z_{ik}^i, z_{ik}^t 分别表示数据 x_{ik} 在第 i 和 t 个任务编码器中映射的隐特征, 即可将式 (1) 中的损失函数简化为

$$\min L(x,z,\theta)=l(x_t,z_t^t)+\lambda H(z_i^t)-\lambda\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\log\frac{\exp(\mathrm{sim}(z_{ik}^i,z_{ik}^t))}{\frac{1}{K}\sum_{j=1}^{K}\exp(\mathrm{sim}(z_{ik}^i,z_{ij}^t))}\right].\tag{5}$$

针对隐向量的熵, 在文献 [27] 中, 借助辅助分布, 给出了近似的上界:

$$H(z)\leqslant\mathbb{E}_p[\log r(z)]\propto\frac{1}{N}\sum_{j=1}^{N}\|z_{\cdot,j}\|_2.\tag{6}$$

鉴于缺少过往任务数据, 选择随机分布 (与原数据均值方差相同的正态分布) 的编码作为当前编码器中的隐变量. 通过约束此编码的范数限制隐向量的熵.

无监督学习的条件下, 异常数据的缺失限制了编码器的性能, 充分挖掘数据之间的一致性能够约束编码器生成适合任务的隐特征, 与此同时, 鉴于式 (5) 中损失函数的形式, 我们对损失的第 1 项 $l(x_t,z_t)$ 也选择互信息损失, 得到最终的损失函数:

$$\min L=-I(x_t,z_t^t)+\lambda\sum_{i=1}^{t-1}[H(z_i^t)-I(z_t^i,z_t^t)].\tag{7}$$

借助 NCE, 应用对比预测编码的互信息下界, 以及式 (6) 中对熵的近似, 可以得到近似优化损失:

$$\begin{aligned}\min L(x,z,\theta) = &-\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\log\frac{\exp(\mathrm{sim}(x_{tk},z_{tk}^{t}))}{\frac{1}{K}\sum_{j=1}^{K}\exp(\mathrm{sim}(x_{tk},z_{tj}^{t}))}\right]\\ &+\lambda\sum_{i=1}^{t-1}\left\{\left\|z_i^t\right\|_2-\frac{1}{K}\mathbb{E}\left[\sum_{k=1}^{K}\log\frac{\exp(\mathrm{sim}(z_{tk}^{i},z_{tk}^{t}))}{\frac{1}{K}\sum_{j=1}^{K}\exp(\mathrm{sim}(z_{tk}^{i},z_{tj}^{t}))}\right]\right\},\end{aligned}\tag{8}$$

其中当前 t 任务的 InfoNCE 损失中使用的 $\mathrm{sim}(x_{tk},z_{tk})$ 参考了文献 [50], 将原始数据在不同的变换下编码得到的隐特征之间的内积作为相似度, 对图像数据有旋转、翻转、裁剪等变换, 对非图像数据可选择仿射变换. $\mathrm{sim}(z_{tk}^{i},z_{tk}^{t})$ 使用内积作为相似度函数, 详见算法 1.

Algorithm 1 Continual unsupervised anomaly detection (CUAD)

Input: Normal sample $\{X_1,\ldots,X_T\}$, batch size K, encoder E_θ, regularization parameter λ, iterations epoch, random distribution X_o.

1: Set $\lambda=0$, $t=1$ for sampled mini-batch:
2: **for** $j=1,\ldots,K$ **do**
3: Select two transforms $\tilde{t}$ and $\hat{t}$;
4: $\tilde{x}_{tj}=\tilde{t}(x_{tj})$, $\hat{x}_{tj}=\hat{t}(x_{tj})$;
5: $\tilde{z}_{tj}^{t}=E_{\theta_t}(\tilde{x}_{tj})$, $\hat{z}_{tj}^{t}=E_{\theta_t}(\hat{x}_{tj})$;
6: $\mathrm{sim}(x_{tj},z_{tj})=(\tilde{z}_{tj}^{t})^{\mathrm{T}}\hat{z}_{tj}^{t}$;
7: $\mathrm{sim}(z_{tk}^{i},z_{tj}^{t})=(z_{tk}^{i})^{\mathrm{T}}z_{tj}^{t}$;
8: $\|z_t\|_2=\frac{1}{2K}\sum_{i}^{K}(\|\tilde{z}_{ti}\|_2+\|\hat{z}_{ti}\|_2)$;
9: $Z^t=E_{\theta_t}(X_o)$;
10: Get loss in (8);
11: Update E_θ, epoch iterations, store Z^t;
12: **end for**
13: **for** $t=2,\ldots,T$ **do**
14: Set $\lambda=\lambda$, do 2∼12 for sampled mini-batch;
15: **end for**

Output: Encoder E_θ.

4 实验与结果

为了验证连续无监督异常检测 CUAD 算法的性能, 本文针对以下关键问题进行了一系列实验: (1) 能否有效地执行无监督异常检测任务? (2) 能否连续地进行无监督异常检测, 对于历史任务中的异常是否都能有效的检测? (3) 能否通过多个任务的学习, 提升在过往任务上的性能?

4.1 数据集及实验设定

实验选取了 4 个普遍选用的数据集, 数据集的基本信息见表 1. MNIST 由 70000 张 28×28 像素的灰度图构成, 训练集包括共 50000 张 0∼9 的图像, 随机设定 5 个类为 5 个检测任务中的正常类, 剩余 5 个类为各自任务中的异常数据, 验证集中也包括了 10 个类共 10000 张图像. 实验时, 我们考虑模型对当前任务中正常样本和异常样本, 以及首个任务中异常数据的检测精度, 以此来回答问题 1 和 2. 在训练时, 每批次 300 个样本, 正则化参数设定为 1/3. 对于精度的度量, 依据文献 [26] 中的做法, 选择使用 ROC 曲线下面积 (AUC) 作为指标.

表 1　实验数据

Table 1　Experimental datasets

Dataset	#classes	#training classes	Data size
MNIST	10	5	28×28×1
CIFAR-FS	100	64	32×32×3
MiniImageNet	100	64	84×84×3
KDDCup99	23	17	41

KDDCup 99[51, 52] 最初为 KDDCup 中用于检测网络入侵的数据集, 包含约 490 万条数据, 共 22 个异常类和一个正常类 Normal, 各类数据的数目各异. 为了构造连续多个无监督异常检测任务, 依据文献 [26] 中的设定: 共 16 个检测任务, 将同属 DOS 攻击的 6 个类作为每个任务的异常类, 其余 16 类的数据依次作为各任务的正常类并用于训练, Normal 类作为初始任务的正常类. 并选择使用 ROC 曲线下面积 (AUC) 作为指标.

MiniImageNet 数据集由 ImageNet 中选取的 100 个类构成, 各类包含 600 张图像, 分别为 84×84 的 RGB 图像, 其中训练集包括 64 个类别, 我们依旧考虑 5 个检测任务, 在训练集和测试集中分别选取 5 个类作为无监督检测任务的正常类与异常类. CIFAR-FS 是将 CIFAR-100 数据集划分生成的, 其中也包括了 100 个类, 64 个类作为训练集, 每个类包括 600 张 32×32 的 RGB 图像, 对于检测任务的设定也和 MiniImageNet 数据集一样. 实验时, 依据文献 [21] 中的标准, 选择了在所有测试任务上的准确率作为指标. 每批次 120 个样本, 正则化参数设定为 0.8.

在整个实验中, 编码器结构包括两层卷积层、两层最大池化层、3 层以 ReLU 作为激活函数的线性层. 使用 Adam 优化器, 学习率为 0.001. 实验中对每次实验都运行 5 次, 并对精度结果取平均.

4.2　对比算法

在 MNIST 和 KDDCup99 数据集上, 我们选取了使用自编码器复现数据的 GR 算法, 以及在连续学习中性能较好的 EWC 算法 [24] 作为对比, 并将不添加过往任务正则化项训练得到的检测结果作为模型性能的下界, 同时将文献 [26] 中, 输入所有检测任务数据得到的结果作为模型性能的上界作为参考. 鉴于在 KDDCup99 数据集上, 性能下界和 EWC 算法的实验结果与其余算法差距较大, 故不在图中画出.

在 MiniImageNet 和 CIFAR-FS 这两个数据集上, 我们选取了借用元学习实现连续异常检测的 ARCADe 算法 [21] 与其中的 ARCADe-H 和 ARCADe-M 模型对比, 同时将正则化参数设为零的测试结果作为基准, 来衡量正则化项的作用.

4.3　实验结果与分析

我们分别给出了 CUAD 与对比算法在不同数据集上的性能对比图. 同时为了统一标准并更全面地展现 CUAD 算法性能, 在表 2 和 3 中给出了具体的 AUC、查准率 (precision)、查全率 (recall) 数值.

对于 MNIST 数据集, 无论是相较于传统的连续学习方法 EWC, 还是使用带标签数据的自编码器 GR 算法, CUAD 在动态任务上始终能保持较高的精度, 在当前任务与学过的首个任务上也都有着出色的表现.

同样的, 在 MiniImageNet 和 CIFAR-FS 数据集上, CUAD 算法也取得了更显著优异的性能, 尤其是相对于没有设置正则化项来适应连续动态场景的 ARCADe 算法, CUAD 在后续任务上有着突出的

表 2 在 MNIST, CIFAR-FS, MiniImageNet 上的实验结果

Table 2 Experiment results on MNIST, CIFAR-FS, MiniImageNet

Datasets		Task1	Task2	Task3	Task4	Task5
MNIST	AUC	0.9936	0.9618	0.9463	0.9341	0.9295
	Precision	0.9991	0.9981	0.9906	0.9934	0.9934
	Recall	0.9883	0.9305	0.9100	0.8881	0.8808
CIFAR-FS	AUC	0.7118	0.7546	0.7882	0.7772	0.7630
	Precision	0.6117	0.7244	0.8257	0.8243	0.7989
	Recall	0.7536	0.8548	0.8318	0.7485	0.7263
MiniImageNet	AUC	0.6897	0.7672	0.7723	0.7493	0.7261
	Precision	0.5534	0.7728	0.8076	0.7567	0.7020
	Recall	0.7128	0.8240	0.7586	0.7328	0.7027

表 3 在 KDDCup99 上的实验结果

Table 3 Experiment results on KDDCup99

Datasets		Task1	Task2	Task3	Task4	Task5	Task6	Task7
KDDCup99	AUC	0.9904	0.9912	0.9928	0.9927	0.9924	0.9923	0.9917
	Precision	0.9928	0.9932	0.9925	1.0000	0.9970	1.0000	1.0000
	Recall	0.9880	0.9892	0.9930	0.9856	0.9880	0.9848	0.9837

Task8	Task9	Task10	Task11	Task12	Task13	Task14	Task15	Task16
0.9917	0.9917	0.9916	0.9915	0.9912	0.9911	0.9911	0.9910	0.9906
1.0000	0.9906	0.9850	1.0000	1.0000	0.9830	0.9850	1.0000	1.0000
0.9837	0.9927	0.9982	0.9832	0.9827	0.9993	0.9972	0.9823	0.9815

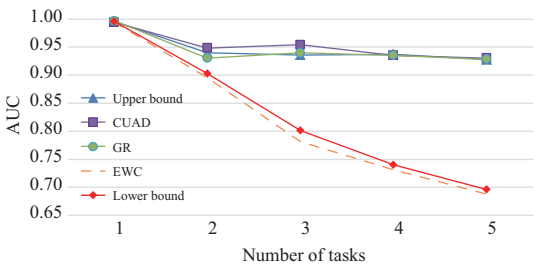

图 1 (网络版彩图) 在 MNIST 上的 AUC 比较

Figure 1 (Color online) AUC on MNIST

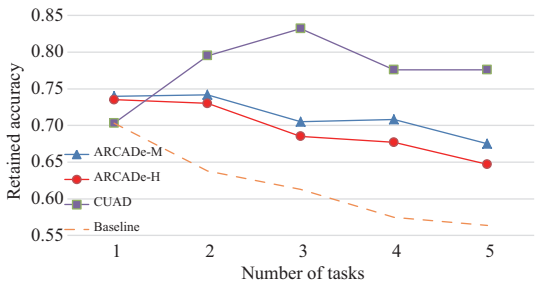

图 2 (网络版彩图) 在 CIFAR-FS 上的精度比较

Figure 2 (Color online) Retained accuracy on CIFAR-FS

表现.

对于非图像数据集的 KDDCup99, CUAD 算法使用了不同于图像增广方式的仿射变换增广, 取得了更为突出的表现. 从第 2 个任务开始就稳定优于 GR 算法的表现.

除此以外, 还有如下发现:

在首个任务上, CUAD 算法的效果均略差于 GR 算法与 ARCADe 算法, 这是因为 GR 算法使用了带标签数据, 有充分的正常与异常样本间的差异信息可用, 而 ARCADe 算法在大量异常任务数据上

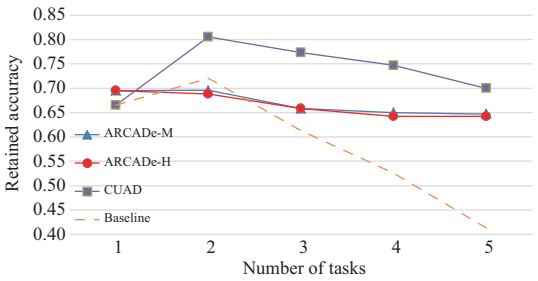

图 3 (网络版彩图) 在 MiniImageNet 上的精度比较

Figure 3 (Color online) Retained accuracy on MiniImageNet

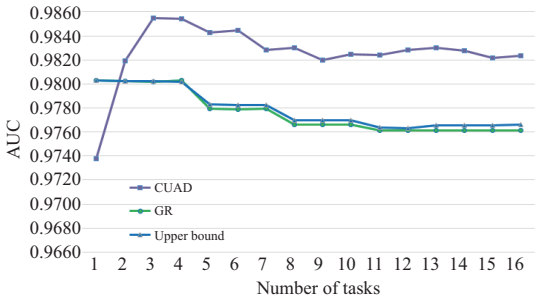

图 4 (网络版彩图) 在 KDDCup99 上的 AUC 比较

Figure 4 (Color online) AUC on KDDCup99

预训练, 其性能提升源于预训练任务, 但此算法未能在连续任务中对模型动态更新, 丢失了新任务中的数据信息. 而 CUAD 算法通过引入互信息, 结合近似优化, 给出了不依赖于异常数据和过往任务数据的损失函数, 减缓了灾难性遗忘问题的同时, 仅需要更少量且无标记的数据, 相较于 GR 与 ARCADe 算法有更广泛的应用场景.

参考无正则化项的对比 (消融) 实验, CUAD 算法不仅能有效地检测当前任务中的异常数据, 也能兼顾到学过的过往任务, 在所有任务上保持较高精度.

随着任务的增加, 模型在 KDDCup99, MiniImageNet 和 CIFAR-FS 数据集上的精度均存在上升现象, 一方面是因为基模型较为简单, 另一方面是模型通过学习后续任务中的正常数据, 得到了更适合这个任务的特征表示. 虽然算法在连续的任务中存在遗忘, 但受益于更多的正常数据和深度模型强大的表示能力, CUAD 从中挖掘数据中蕴含的深层表示信息, 刻画出更能准确表征正常样本的特征, 即回答了本节提出的问题 3.

连续学习的目的不仅在保证新任务性能的基础上不遗忘过往任务知识, 还希望能够通过后续任务的学习巩固提升过往任务的性能, 我们可以将其视为知识在多个任务之间的传递.

5 总结与展望

本文考虑动态变化的连续无监督异常检测问题, 基于互信息, 通过对损失函数的分解优化, 设计出了适合动态场景的损失函数, 并提出了一种新的连续检测异常的无监督算法 CUAD. 实验表明在多个数据集上, 本文具有比现有方法更优的性能.

未来的改进方向是结合元学习为 CUAD 构建一个更优秀的初始基模型, 并进一步利用检测出的异常数据. 将这些数据用于训练可在动态环境下弥补异常数据的缺失. 同时, 对于异常数据的具体定位和分类均为后续改进的研究方向.

此外, 在实际应用场景中, 如何对异构数据进行处理, 使其能够适应连续无监督异常检测模型也是值得探索的问题. CUAD 算法主要研究的是维度一致的数据, 可以考虑调整模型结构, 并结合更多变换技术.

参考文献

1 Chalapathy R, Menon A K, Chawla S. Anomaly detection using one-class neural networks. 2018. ArXiv:1802.06360

2 Wei Q, Ren Y H, Hou R, et al. Anomaly detection for medical images based on a one-class classification. In: Proceedings of SPIE, 2018
3 Hejazi M, Singh Y P. One-class support vector machines approach to anomaly detection. Appl Artif Intell, 2013, 27: 351–366
4 García-Teodoro P, Díaz-Verdejo J, Maciá-Fernández G, et al. Anomaly-based network intrusion detection: techniques, systems and challenges. Comput Secur, 2009, 28: 18–28
5 Malhotra P, Ramakrishnan A, Anand G, et al. LSTM-based encoder-decoder for multi-sensor anomaly detection. 2016. ArXiv:1607.00148
6 Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv, 2009, 41: 1–58
7 Park H, Noh J, Ham B. Learning memory-guided normality for anomaly detection. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020. 14372–14381
8 Ruff L, Vandermeulen R, Goernitz N, et al. Deep one-class classification. In: Proceedings of International Conference on Machine Learning, 2018. 4393–4402
9 Smola A, Schölkopf B. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge: MIT Press, 2002
10 Eskin E, Arnold A, Prerau M, et al. A geometric framework for unsupervised anomaly detection. In: Applications of Data Mining in Computer Security. Berlin: Springer, 2002. 77–101
11 Bengio Y, LeCun Y. Scaling learning algorithms towards AI. In: Proceedings of Large-Scale Kernel Machines, 2007
12 Socher R, Bengio Y, Manning C D. Deep learning for NLP (without magic) references. 2012. https://nlp.stanford.edu/~socherr/DeepLearning-ACL2012-tutorial.pdf
13 Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis, 2015, 115: 211–252
14 Chen J X. The evolution of computing: alphaGo. Comput Sci Eng, 2016, 18: 4–7
15 Zhou C, Paffenroth R C. Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017. 665–674
16 Chalapathy R, Menon A K, Chawla S. Robust, deep and inductive anomaly detection. In: Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2017. 36–51
17 Andrews J T A, Morton E J, Griffin L D. Detecting anomalous data using auto-encoders. Int J Mach Learn Comput, 2016, 6: 21
18 Erfani S M, Rajasegarar S, Karunasekera S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recogn, 2016, 58: 121–134
19 Tax D M J, Duin R P W. Support vector data description. Mach Learn, 2004, 54: 45–66
20 Pan K, Palensky P, Esfahani P M. From static to dynamic anomaly detection with application to power system cyber security. IEEE Trans Power Syst, 2020, 35: 1584–1596
21 Frikha A, Krompaß D, Tresp V. ARCADe: a rapid continual anomaly detector. 2020. ArXiv:2008.04042
22 Delange M, Aljundi R, Masana M, et al. A continual learning survey: defying forgetting in classification tasks. IEEE Trans Pattern Anal Mach Intell, 2021. doi: 10.1109/TPAMI.2021.3057446
23 French R. Catastrophic forgetting in connectionist networks. Trends Cogn Sci, 1999, 3: 128–135
24 Kirkpatrick J, Pascanu R, Rabinowitz N, et al. Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci USA, 2017, 114: 3521–3526
25 Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of International Conference on Machine Learning, 2017. 1126–1135
26 Wiewel F, Yang B. Continual learning for anomaly detection with variational autoencoder. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019. 3837–3841
27 Ye F, Zheng H J, Huang C Q, et al. Deep unsupervised image anomaly detection: an information theoretic framework. 2020. ArXiv:2012.04837
28 Candés E J, Li X D, Ma Y, et al. Robust principal component analysis? J ACM, 2011, 58: 1–37
29 Li D, Chen D C, Goh J, et al. Anomaly detection with generative adversarial networks for multivariate time series. 2018. ArXiv:1809.04758
30 Parzen E. On estimation of a probability density function and mode. Ann Math Statist, 1962, 33: 1065–1076

31 Yang B, Fu X, Sidiropoulos N D, et al. Towards k-means-friendly spaces: simultaneous deep learning and clustering. In: Proceedings of International Conference on Machine Learning, 2017. 3861–3870

32 Liu F T, Ting K M, Zhou Z H. Isolation-based anomaly detection. ACM Trans Knowl Discov Data, 2012, 6: 1–39

33 Liu F T, Ting K M, Zhou Z H. Isolation forest. In: Proceedings of the 8th IEEE International Conference on Data Mining, 2008. 413–422

34 Schlegl T, Seeböck P, Waldstein S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Proceedings of International Conference on Information Processing in Medical Imaging, 2017. 146–157

35 Andrews J, Tanay T, Morton E J, et al. Transfer representation-learning for anomaly detection. In: Proceedings of the 33rd International Conference on Machine Learning, 2016

36 Doshi K, Yilmaz Y. Continual learning for anomaly detection in surveillance videos. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020. 254–255

37 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

38 Li C L, Sohn K, Yoon J, et al. CutPaste: self-supervised learning for anomaly detection and localization. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021. 9664–9674

39 Rebuffi S A, Kolesnikov A, Sperl G, et al. ICARL: incremental classifier and representation learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017. 2001–2010

40 Chaudhry A, Ranzato M A, Rohrbach M, et al. Efficient lifelong learning with A-GEM. 2018. ArXiv:1812.00420

41 Lopez-Paz D, Ranzato M A. Gradient episodic memory for continual learning. In: Proceedings of the 31st Conference on Neural Information Processing System, 2017. 6470–6479

42 Zenke F, Poole B, Ganguli S. Continual learning through synaptic intelligence. In: Proceedings of International Conference on Machine Learning, 2017. 3987–3995

43 Li Z Z, Hoiem D. Learning without forgetting. IEEE Trans Pattern Anal Mach Intell, 2017, 40: 2935–2947

44 Mallya A, Lazebnik S. Packnet: adding multiple tasks to a single network by iterative pruning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 7765–7773

45 Rusu A A, Rabinowitz N C, Desjardins G, et al. Progressive neural networks. 2016. ArXiv:1606.04671

46 Chen P H, Wei W, Hsieh C J, et al. Overcoming catastrophic forgetting by Bayesian generative regularization. In: Proceedings of International Conference on Machine Learning, 2021. 1760–1770

47 Hou S, Pan X, Loy C C, et al. Lifelong learning via progressive distillation and retrospection. In: Proceedings of European Conference on Computer Vision (ECCV), 2018. 437–452

48 Qu H X, Rahmani H, Xu L, et al. Recent advances of continual learning in computer vision: an overview. 2021. ArXiv:2109.11369

49 Oord A, Li Y Z, Vinyals O. Representation learning with contrastive predictive coding. 2018. ArXiv:1807.03748

50 Bergman L, Yedid H. Classification-based anomaly detection for general data. 2020. ArXiv:2005.02359

51 Tavallaee M, Bagheri E, Lu W, et al. A detailed analysis of the KDD CUP 99 data set. In: Proceedings of IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009. 1–6

52 Stolfo S J, Fan W, Lee W, et al. Cost-based modeling for fraud and intrusion detection: results from the JAM project. In: Proceedings DARPA Information Survivability Conference and Exposition, 2000. 130–144

Continual unsupervised anomaly detection

Yiming NI[1,2] & Songcan CHEN[1,2*]

1. *College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing* 210016*, China*;
2. *MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing* 210016*, China*
* Corresponding author. E-mail: s.chen@nuaa.edu.cn

Abstract Unsupervised anomaly detection (UAD) involves detecting unseen data that differs from expected or normal patterns. UAD has gained extensive attention in dynamic scenarios because its learning process is independent of rare abnormal samples. In reality, target tasks change dynamically with time, so a model must solve different UAD problems continuously and effectively, detecting all kinds of previous anomalies with only normal data in the current task. Thus, we propose a novel continual UAD (CUAD) algorithm based on mutual information. Specifically, the original objective function needs previous and abnormal data, which are missing. To deal with this problem, we introduce an objective function and then approximate and optimize it based on information theory. On the basis of this, a deep encoder is used to continuously detect various anomalies while efficiently alleviating catastrophic forgetting caused by continual learning. Experiments on several datasets demonstrate that the proposed model outperforms state-of-the-art methods.

Keywords anomaly detection, unsupervised, catastrophic forgetting, continual learning, information theory