SwitchTab: 切换自动编码器是有效的表格学习器

吴菁*, 陈随耀*, 赵琪, Renat Sergazinov、陈力、刘胜杰、赵崇超、谢天培、郭汉清、吉成、 丹尼尔・科乔瓦,哈坎・布伦泽尔

> 亚马逊买家风险防范 4575 La Jolla Village Dr 圣地亚 哥,加利福尼亚州 92122 美国

fiingwua, suiyaoc, qqzhao, renserg, chenlii, zycilsi, zchongch, lukexie, hanqiguo, cjiamzn, cociorva, brunzell/@amazon.com

抽象的

自监督表征学习方法在计算机视觉和自然语言处理 (NLP) 中取得了重大成功,其中数据样本表现出明确的空间或语 义依赖关系。然而,由于数据样本之间的依赖关系不太明 显,将这些方法应用于表格数据具有挑战性。在本文中, 我们通过引入 SwitchTab 来解决这一限制, SwitchTab 是 一种专门设计用于捕获表格数据中潜在依赖关系的新型自 监督方法。SwitchTab 利用非对称编码器-解码器框架来解 耦数据对之间的相互和显着特征,从而产生更具代表性的 嵌入。这些嵌入反过来有助于更好的决策边界并导致下游 任务的结果得到改善。为了验证 SwitchTab 的有效性,我 们在涉及表格数据的各个领域进行了广泛的实验。结果展 示了通过微调在端到端预测任务中的卓越性能。此外,我 们证明预训练的显着嵌入可以用作即插即用功能来增强各 种传统分类方法(例如 Logistic 回归、XGBoost 等)的性 能。最后,我们强调了 SwitchTab 通过在潜在空间中可视 化解耦的相互特征和显著特征来创建可解释表示的能力。

介绍

虽然表征学习 (Bengio、Courville 和 Vincent 2013) 在计算机视觉 (CV) 和自然语言处理 (NLP) 领域取得了显著进步,但表格数据在现实世界的应用和关键行业中无处不在,例如医疗保健 (Qayyum et al. 2020; Chen et al. 2017, 2019)、制造业 (Borisov et al. 2022; Chen, Lu, and Li 2017; Wang et al. 2023a; Chen et al. 2020a)、农业 (Liakos et al. 2018; Wu et al. 2022; Tao et al. 2022) 和各种工程领域 (Chen et al. 2018; Zhu et al. 2018; Chen 2020; Wang et al. 2023b),它尚未充分受益于其变革力量,并且仍然相对落后尚未探索。表格数据集带来的独特挑战源于其固有的异质性,

*这些作者做出了同等的贡献。 版权所有©2024,人工智能促进协会 情报(www.aaai.org)。保留所有权利。

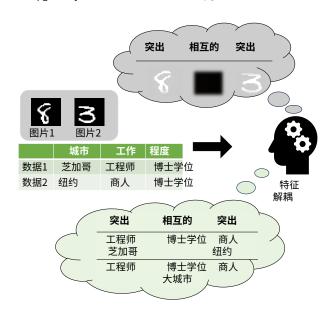


图 1: 给定一对图像,由于结构良好的空间关系,人们可以轻松区分突出的数字和共同的背景。然而,区分一对表格样本却变得具有挑战性。例如,特征 City 可能在数据点"芝加哥"和"纽约"之间显著,但仍然共享一些潜在的相互信息(例如大城市),这使得解耦变得具有挑战性。请注意,此解耦过程仅用于说明。在实现中,所有解耦样本都在特征空间中计算。

缺乏图像中的明确空间关系(例如,相似的背景和不同的字符)或语言中的语义依赖关系。表格数据通常包含数字和分类的冗余特征,表现出各种离散和连续分布

(Grinsztajn、Oyallon 和 Varoquaux 2022)。这些特征可以相互依赖或完全独立,这使得表征学习模型难以捕获关键的潜在特征,从而无法在不同样本中进行有效决策或准确预测。

在比较数据样本时,相互特征包括

信息突显了共同特征,而显著特征则强调了独特属性,以区分一个样本与其他样本。对于图像数据,背景像素的强度形成图像间共享的相互特征,而亮像素和暗像素的相对位置形成显著特征,这些特征在不同形状或物体的图像之间可能会有显著差异。如图 1 所示,在 MNIST(Xiao、Rasul 和 Vollgraf 2017)中,将数字与背景分离相对简单,使用数字作为分类的显著特征。然而,表格数据的区分往往不那么明显。例如,当数据点"芝加哥"和"纽约"的字数不同时,像城市这样的特征可以被认为是显著的。尽管如此,当从语义上考虑城市的大小时,城市特征可以共享相互信息。因此,设置分类的决策边界变得更加复杂。

为了应对这些挑战,我们的核心见解是让表征模型能够明确区分特征空间中的相互信息和显著信息,我们将其定义为解耦过程。我们坚信,操纵特征空间可以减少噪音并获得更多的代表性,而不是仅仅依赖原始数据空间,从而将其他领域的表征学习的成功应用于表格数据。

在本文中,我们介绍了 SwitchTab,这是一个用于表格数据表示学习的优雅而有效的生成预训练框架。 SwitchTab 的核心是一种非对称编码器-解码器结构,并增加了便于信息解耦的自定义投影仪。该过程从将每个数据样本编码为一般嵌入开始,然后进一步投影为显着嵌入和相互嵌入。SwitchTab 的与众不同之处在于在解码过程中刻意在不同数据样本之间交换显着嵌入和相互嵌入。这种创新方法不仅允许模型从编码器获取更结构化的嵌入,还可以明确提取和表示显着信息和相互信息。SwitchTab的另一个优点是它的多功能性,可以在自监督方式下进行有效训练。这种适应性确保 SwitchTab 在各种训练场景中表现良好,无论是否有标记数据。

我们的贡献可以总结如下:

- ·我们提出了 SwitchTab,这是一种新颖的自监督学习框架,用于分离数据样本中的显着嵌入和相互嵌入。 据我们所知,这是首次尝试探索和明确提取表格数据的可分离和有组织的嵌入。
- ·通过对 SwitchTab 的预训练编码器进行微调,我们证明我们的方法在广泛的数据集和基准上取得了有竞争力的结果。
- ·提取的显著嵌入可以作为即插即用功能来增强各种传统预测模型的性能,例如 XGBoost。
- ·我们将从 SwitchTab 中学到的结构化嵌入可视化,并 强调相互信息和显著信息之间的区别,从而增强所提 框架的可解释性。

相关工作

表格数据学习和预测模型

传统模型。对于表格数据分类和回归任务,已经开发了各种机器学习方法。对于线性关系建模,逻辑回归 (LR) (Wright 1995) 和广义线性模型 (GLM) (Hastie 和 Pregibon 2017) 是首选。基于树的模型包括决策树 (DT) (Breiman 2017) 和基于 DT 的各种集成方法,例如 XG-Boost (Chen 和 Guestrin 2016)、随机森林 (Breiman 2001)、CatBoost (Prokhorenkova 等人 2018) 和 Light-GBM (Ke 等人 2017),这些方法在行业中被广泛用于建模复杂的非线性关系、提高可解释性和处理各种特征类型,如空值或分类特征。

深度学习模型。最近的研究趋势旨在将深度学习模型应用 于表格数据领域。已经引入了各种神经架构来提高表格数 据的性能。有几个主要类别(Borisov等人,2022年; Gorishniy 等人,2021年),包括1)使用神经网络的监 督方法(例如,ResNet(He 等人,2016 年)、SNN (Klambauer 等人, 2017年)、AutoInt(Song 等人, 2019年)、DCN V2(Wang 等人, 2021年));2)将 决策树与神经网络集成以进行端到端训练的混合方法(例 如, NODE (Popov、Morozov 和 Babenko, 2019 年)、GrowNet(Badirli 等人,2020年)、TabNN(Ke 等人,2018年)、DeepGBM(Ke 等人,2019年)) 3) 基于 Transformer 的方法,从跨特征和数据样本的注 意力中进行学习(例如 TabNet(Arik 和 Pfister 2021) 、Tab-Transformer (Huang 等人 2020) 、FT-Transformer (Gorishniy 等人 2021)); 4) 表示学习 方法,这些方法具有新兴的焦点并与我们提出的工作范围 相一致,通过自监督和半监督学习实现有效的信息提取 (例如 VIME (Yoon 等人 2020) 、SCARF (Bahri 等人 2021) 、SAINT(Somepalli 等人 2021)) 和 Recontab (Chen 等人 2023)。

自监督表征学习

深度表示学习方法已引入计算机视觉和遥感领域,利用自 监督学习方法(Kolesnikov、Zhai 和 Beyer 2019; Ericsson 等人 2022; Li、Guo 和 Schuurmans 2015; Wu、Hobbs 和 Hovakimyan 2023; Manas 等人 2021; Wu、Hovakimyan 和 Hobbs 2023)。这些方法可分为两 个分支。第一个分支主要关注具有各种数据增强方案的对 比学习框架。更具体地说,模型依赖于动量更新策略(He 等人, 2020; Wu 等人, 2023; Chen 等人, 2020c; Wu、Hovakimyan 和 Hobbs, 2023) 、大批量(Chen 等人, 2020b)、停止梯度操作(Chen 和 He, 2021), 或训练在线网络以预测目标网络的输出(Grill等人, 2020)。这些想法也已应用于表格数据领域。该领域的一 个代表性工作是 SCARF(Bahri 等人,2021),它采用了 SimCLR(Chen 等人,2020b)的思想,使用特征损坏作 为数据增强方法对编码器进行预训练。另一项工作是 SAINT(Somepalli 等人,2021),它也源于对比学习

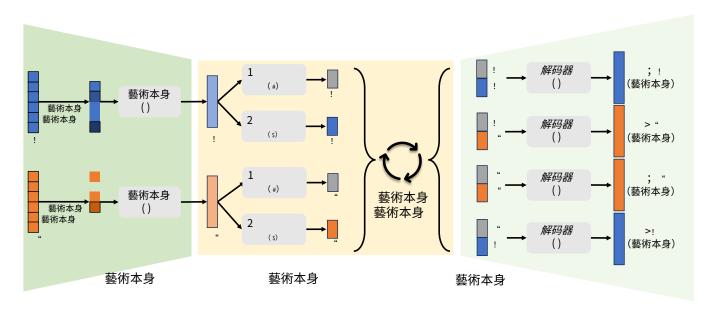


图 2:本文提出的自监督学习框架的框图。(1)两个不同的样本 X_1 和 X_2 被随机破坏并编码成特征向量 E_1 和 E_2 通过编码器 F_2 (2)特征向量 E_1 和 E_2 通过两个不同的投影仪分解为相互和显著的特征 E_3 来和 E_4 (3)解码器将相互特征和显著特征进行组合和重构 E_4 4中显著特征主导样本类型,而相互特征提供可在两个样本之间切换的共同信息。

框架并计算列式和行式注意力。第二个分支基于自动编码器等生成模型(Kingma 和 Welling 2013)。具体来说,蒙版自动编码器 (MAE)(He 等人,2022 年)具有非对称编码器-解码器架构,用于从图像中学习嵌入。该框架还能够捕获时空信息(Feichtenhofer 等人,2022 年),并且可以扩展到 3D 空间(Jiang 等人,2022 年)

和多个尺度(Reed 等人,2022 年)。类似的掩蔽策略义 在 NLP 中被广泛使用(Devlin 等以及 2018年)作为 布尔数据(Arik和Pfister 2021; Huang等人 2020; 阴 等al. 2020)。与 MAE ar 数据类似 在做 ta 的主要内容 布衡工作是 VIME(Yoon 等人, VIME 柯 恩科(020)。在每个估计器之后,将使用两个 估计量 **A**特征空间中的每个样本分解为特 重新分配 呃证,以分别重建二进制掩码和样 原本的 带解码 d 本。我们工作的关键区别在于我们参考 吐温 VIM 埃利用 不对称 我们在 2023 年的预训练中改进了编码器-解码 (陈等人 器架构,并引入了一种切换机制,鼓励编码器 我强烈 牛成更具结构性的嵌入。 红色和代表

特征D耦鉄

Zhang 等人 2020 年提出的一种新方法,其中每个维度都可以捕捉一个语义上有意义的变异因素的变化,同时相对不受其他因素变化的影响。最近的研究还探索了如何捕捉不同变异因素之间的依赖关系和关系,以增强潜在表示(Sønderby 等人 2016 年;Tschannen、Bachem 和Lucic 2018 年)。更进一步,Abid 和

Zou 2019),该研究采用了对比分析的原则, ples,明确地按显着和相互信息对潜在特征进行分类,并 增强了显着特征。 (Park et al. 2020) 的交换自动编码器明 确地将图像分解为结构和纹理嵌入,

被交换用于图像生成。最近一些针对表格数据表示学习的 研究也表明

量化样本间关系的好处。关系自动编码器 (RAE) (Meng et al. 2017) 同时考虑了数据特征和关系,以生成具有较低重构损失和更好预测的更稳健的特征。

ter 在下游任务中的表现。(Kossen 等人,2021 年; Somepalli 等人,2021 年)分享了类似的想法,考虑数 据样本之间的自注意力。我们扩展了 cVAE 的想法 并将自动编码器交换到表格数据域,理由是两个数据样本 通过潜在的样本间关系共享相互和显著的信息。显著信息 对于下游任务至关重要

涉及决策边界,而相互信息对于数据重建仍然是必要的。 尽我们所能

知识,我们第一个用明确和富有表现力的特征解耦架构对 表格数据进行建模,以增强表示学习性能。

算法 1: 使用 SwitchTab 进行自监督学习

要求:未标记数据 \subseteq R*,批次大小 \bigcirc 、编码器 F,亲相互信息喷射器 \bigcirc 0*、投影仪用于突出信息 \bigcirc 0、解码器 \bigcirc 0,均方误差 MSE,特征 con-连锁 \bigcirc 0。

3: 数据编码: $\Delta = F(X_{1,2}), \mathcal{L}_{2,2} = F(X_{2,2}), \text{为了} \mathcal{A} \in [Z]$

ス度: Γ (Λ1我), 定我一Γ (Λ2我), 万 [3 Λ・ T+ 台上 6刀 ‡甲・

4: 功能解耦:

定义如下: S_2 E=B E=B

5: 数据重建:

(1) 让恢复对定义为:

 $X_{\mathbb{H}}=d (\mathcal{K}_{1\mathfrak{H}}\oplus S_{1\mathfrak{H}}), X_{\mathfrak{H}}=d (\mathcal{K}_{2\mathfrak{H}}\oplus S_{2\mathfrak{H}})$

(2) 设交换对定义为:

 $X_{\mathbb{H}}=d (\mathcal{K}_{1\mathcal{H}} \oplus S_{1\mathcal{H}}), X_{\mathcal{H}}^2=d (\mathcal{K}_{1\mathcal{H}} \oplus S_{2\mathcal{H}})$

6: 定义重建损失 大号侦察

均方误差(X銀,X銀)+均方误差(X2 X/3)+均方误差(X1,3,X我)+均方误差(X2 X4)

更新编码器户、投影仪页*和页。和解码器d最小化大号侦察使用 RMSProp。

8: 结束于

方法

在本节中,我们将介绍 SwitchTab,这是我们用于表格数据表示学习和特征解耦的综合方法。首先,我们概述了特征损坏的过程。然后,在第二小节中,我们深入研究了自监督学习的复杂性,包括数据编码、特征解耦和数据重建。第三小节阐明了我们的带标签预训练学习方法。最后,我们说明如何利用预训练的编码器和嵌入来改进下游任务。

功能损坏

基于生成的表征学习依赖于数据增强来学习下游任务的稳健嵌入。在不同的方法中,特征损坏(Yoon等人,2020年;Bahri等人,2021年)是最有前途的方法之一。在本文中,我们还利用这种方法来提高模型性能。对于一个表格数据 $X_{\mathcal{R}}$ 来自原始数据集 \subseteq R $_{\mathcal{R}}$,我们定义它的 $_{\mathcal{K}}$ 等个特征为 $_{\mathcal{K}}$,IE, $_{\mathcal{K}}$ 是样本的索引。对于每个样本,我们随机选择 $_{\mathcal{K}}$ 特征 $_{\mathcal{K}}$ 为能并将其替换为损坏的功能 $_{\mathcal{K}}$. 反对

克里特利, $\{C \sim X_{\mathcal{R}}\}$,在哪里 \mathcal{R} 是均匀分布 $X X_{\mathcal{R}, n} = X_{\mathcal{R}, n} : X_{\mathcal{R}} \in X_{\mathcal{R}}$.

自监督学习

SwitchTab 的自监督学习旨在从未标记数据中学习信息表示(算法 1),如图 2 所示。对于两个数据样本中的每一个, λ 和 λ ,我们应用特征损坏来获取损坏的数据。我们使用编码器对它们进行编码,F,结果

在两个特征向量中, \mathcal{L}_1 和 \mathcal{L}_2 。重要的是,我们使用两种类型的投影仪将这两个特征向量分离, \mathcal{D}_* 和 \mathcal{D}_s ,分别提取数据样本之间可切换的互信息和每个数据样本独有的显著信息。通过这个解耦过程,我们得到了显著特征向量, \mathcal{L}_1

和 \mathfrak{L} 以及相互特征向量, \mathfrak{L} 和 \mathfrak{L} 2,为了 \mathfrak{L} 1 和 \mathfrak{L} 2,分别。

值得注意的是,相互的特征应该在两个样本之间共享和切换。换句话说,连接的特征向量 $s_1 \oplus \mathcal{H}_1$ 应该没有表现出明显差异相比 $s_1 \oplus \mathcal{H}_2$ 。因此,预计不仅解码的数据 X_1 (恢复) $s_1 \oplus \mathcal{H}_1$ 高度相似 X_1 ,还有解码后的数据 X_1

(切换)来自连接的特征向量s₁ *⊕ 米*₂ 应该表现出相当的相似性。同样,我们预计两者*X*₂ (恢复)和*X*₂ (切换)类似于*X*因此,我们定义损失函数*大号自己 大号侦察*重建损失为:

大号侦察=
$$\frac{1}{\cancel{K}}$$
 $\frac{1}{\cancel{K}}$ $\frac{1}{\cancel{K}$

使用标签进行预训练

我们利用标记数据进一步改进了预训练过程,如图 3 所示。通过引入标签,我们对编码的嵌入施加了额外的约束 是1和是2用于标签预测并计算预测损失(以分类损失为例 大号分类通过上下文)。具体来说,是1和是2被输入到从嵌入空间映射到标签空间的同一个多层感知器 (MLP)。在优化阶段,我们将预测损失与大号侦察更新框架中的参数。正式地,我们定义损失函数大号全部的对于两个样本从和处如下:

大号
$$_{2}$$
部的-大号 $_{6}$ 寮+ α^* 大号 $_{6}$ 类, (2)

在哪里 α 用于平衡分类损失和重构损失,默认设置为 1。为了说明,分类任务的交叉熵损失定义为:

大号分类——(是1日志(
$$\hat{y}_1$$
)+是2日志(\hat{y}_2)), (3)

在哪里 \hat{p}_1 和 \hat{p}_2 是预测标签,即 \hat{p}_1 =MLP(\mathcal{E}_1)和 \hat{p}_2 =MLP(\mathcal{E}_2 对于回归任务,根均方误差(RMSE)将取代交叉熵损失。

下游微调

根据已建立的表征学习范式(He 等人,2020 年;Chen 等人,2020c、b;Bahri 等人,2021 年),我们使用完整的标记数据集对 SwitchTab 中的预训练编码器进行端到端微调。具体来说,我们将编码器 F另外

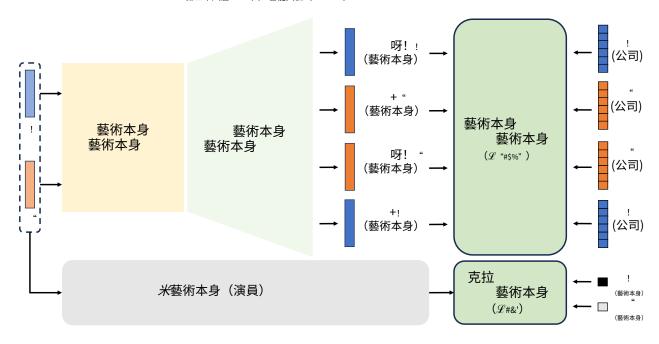


图 3: 带标签的预训练框架框图。(1)监督学习: 潜在特征向量 *是*1和 *是*2通过多层感知器(MLP)预测标签。根据预测标签和真实标签计算交叉熵损失。(2)自监督学习: 重建(恢复和切换)数据和原始编码数据用于计算均方误差(MSE)。

线性层,解锁其所有参数并使其适应下游监督任务。

A利用 o 优势的另一种途径 你的框架工作 在于利用显着特征向量 s 播放嵌入。通过连作为插件和-接 s 及其 re 向量 X,我们构建了丰富的数据样 原来的特色本 X 连接— X ④ s 该方法有效地识别了数据中的不向量德诺同特征,同时建立了明确的决策界限,我们 伊利高光 预计在使用时,c 任务会有明显的增强 X 连接 作 ch 设施它吃了XGBoost等 tr el 的输入。 作为结果 分类化 附加模式

实验与结果

在本节介绍了我们在各种数据集上进行的 紧密的经验SwitchTab 有效性测试的结果。本节分为五示威埃 个部分。在第一部分中,我们提供了有关实融入二验的初步信息,包括数据集、测试、模型架信息化构和训练细节,以确保透明度和可重复性。数据预进程 瞄准 留 en

在 第二部分,我们评估表现 我们的从两个不同的角度展示了我们提出的方法。首先,我们使用来自(Gorishniy et al. 2021)的标准基准和其他数据集将 SwitchTab 与主流深度学习和传统模型进行比较,以建立对 SwitchTab 的更全面的性能评估。其次,我们通过演示展示了 SwitchTab 的多功能性。

开始在各种传统模型中使用即插即用的嵌入显着特征,包 括

实验准备

数据集。我们首先根据标准基准对性能进行评的切换标签估,这些基准来自(Gorishniy et a cretely,数2021). 反对据集包括: California Housin and Barry G ck)(佩斯)1997)、Adult (AD) (Kohavi et al. 1 (HE) 996)海伦娜(Guyon et al. 2019b)、Jannis (JA) (Guyon 等2019b),Higgs (HI) (Baldi, Sadowski, and Whiteson (AL) (Geusebroek, Burghouts, and Smeulde 章频05),Epsilon (EP)(Yuan、Ho和Lin 2011)、年份(YE)(Bertin-Mahieux等人 2011)、Covertype (CO)(黑色)阿尔德迪恩1999)、雅虎(YA)(Chapelle and Chang 201(MI)1)微软(Qin and Liu 2013)。

除了标准基准之外,还有最近的研究(So 所**赐**一组2021)使用的许多流行数据集,包括 Bank 我 palli 等人(BK)(Moro、Cortez、Blastchar (BC) d丽塔 2014)(Ouk、Dada 和 Kang 2018)心律失常(AT)(Liu、Ting 和 Zhou 2008;Ouk、Dada 和 Kang 2018)、Arcene (AR)(亚松森和纽曼 2007)、店铺pers (SH)(Sakar et al. 2019)、Volkert (VO) (**G**yon 等人 2019a)和 MNIST (MN)(Xiao、Rasul 和 V) 奥**尔**17年)。

数据集的预处理。我们使用后向差分编码器 特**權**论特征(dawala 和 Pai 2017)来表示图像。关于 (**如**ar, Par-缺失数据 噢,我们丢弃所有样本都缺少任何特征。对于重新针对缺失值,我们采用基于特征类型的插补策略。数值特征被插补

数据集大小 特征尺寸	48842 65196 83733 98050 108000 500000 51 14275428128	.8012 20640 51534 2000 54	5 709877 120019 8 90	699 136
方法/数据集	广告 ↑ 他 ↑ JA ↑ 你好 ↑ 艾尔 ↑	EP	加州↓叶↓	呀↓ MI↓
塔格网 信噪比 自动输入 PERFENDENTS DCN2 节点 残差网络 FT变压器	0.850 0.378 0.723 0.719 0.954 0.854 0.373 0.719 0.722 0.954 0.859 0.372 0.721 0.725 0.945 0.852 0.383 0.723 0.723 0.954 0. 853 0.385 0.723 0.723 0.955 0.858 0.359 0.726 0.726 0.918 0.8540.3960.727 0.727 0.963 0.859 0.391 0.729 0.729 0.960	0.8896 0.957 0.8975 0.961 0.8949 0.934 0.8977 0.962 0.8977 0.965 0.8958 0.985 0.8969 0.964 0.8982 0.970	0.510 8.909 0.493 8.895 0.474 8.882 0.499 8.853 0.484 8.890 0.464 8.784 0.486 8.846 0.459 8.855	0.761 0.751 0.768 0.750 0.757 0.747 0.757 0.749 0.753 0.745 0.757 0.748
XGBoost CatBoost SwitchTab (自支持) 切换标签	0.874 0.377 0.724 0.728 0.924 0.873 0.388 0.727 0.729 0.948 0.867 0.387 0.726 0.724 0.942 0.881 0.389 0.731 0.733 0.951	0.8799 0.964 0.8893 0.950 0.8928 0.971 0.8987 0.989	0.431 8.819 0.423 8.837 0.452 8.857 0.442 8.822	0.740 0.743 0.755 0.751

使用平均值,而分类特征则用数据集中最常见的类别填充。此外,我们使用 Min-Max 缩放器缩放数据集来确保一致性。在处理基于图像的数据时,我们将它们展平为向量,从而将它们视为表格数据,遵循先前工作中建立的方法(Yoon 等人,2020 年;Somepalli 等人,2021年)。

模型架构。对于特征损坏,我们为每个样本均匀采样一个特征子集,以固定损坏率为 0.3 生成损坏视图。对于编码器 F,我们采用带有两个头的三层变压器。编码器的输入和输出大小始终与输入的特征大小对齐。两个投影仪 页。

和 页**由一个线性层和随后的 S 型激活函数组成。此外,解码器 d仍然是具有 sigmoid 激活函数的单层网络。在带标签的预训练阶段,我们引入了一个额外的单层网络进行预测。在下游微调阶段,我们在编码器后附加了一个线性层 F以适应分类或回归任务。

培训详情。重要的是,我们在 SwitchTab 的整个评估过程中保持一致的设置。虽然通过进一步探索超参数可能会获得进一步的收益,但我们有意不这样做,以确保所提出的方法可以轻松地推广到不同的数据集和领域。对于所有预训练,我们对所有模型进行 1000 次训练,默认批次大小为 128。我们使用 RMSprop 优化器(Hinton、Srivastava 和 Swersky 2012),初始学习率设置为0.0003。在微调阶段,我们将最大时期设置为 200。使用Adam 优化器,学习率为 0.001。

先前基准测试的结果

我们对来自之前基准测试的 11 个数据集上的 SwitchTab 与不同方法进行了全面的性能比较,如表 1 所示。为了确保公平和直接的比较,我们报告了分类准确率。

分类任务,遵循先前研究中使用的指标。 值得注意的是,我们根据既定范式(Kolesnikov、Zhai 和 Beyer 2019)对结果进行了细致的微调。分析结果后,我们发现SwitchTab 在大多数分类任务中始终实现最佳或接近最佳的性能。这些结果强调了 SwitchTab 在分类场景的表示学习中的有效性和优越性。然而,在回归任务中,我们观察到 XGBoost 或 CatBoost 等传统方法仍然占主导地位并取得最佳效果。尽管如此,SwitchTab 仍然具有很强的竞争力,并且在这些回归场景中优于各种深度学习方法。我们报告了 10 个随机种子的平均结果。

其他公共数据集上的结果

除了之前的基准测试之外,我们继续在其他公共数据集上进行性能比较,并在表 2 中总结了结果。结果包括使用传统模型和较新的深度学习技术进行的评估。在大多数情况下,SwitchTab 表现出显着的改进,超越了所有基线方法,并巩固了其在不同数据集和场景中的优势。然而,必须承认,在数据集 AT 上,与基线相比,SwitchTab 取得了次优结果。这一观察结果与先前的研究结论一致,即表格领域带来了独特的挑战,其中没有一种方法普遍占主导地位(Gorishniy等人,2021年)。然而,这一结果值得进一步研究,以辨别导致这种性能变化的具体因素。

即插即用嵌入

如前所述,SwitchTab 擅长有效提取显著特征,这些特征可能会显著影响分类任务的决策边界。在即插即用设置中,我们的实验结果表明,这些显著特征在集成时具有巨大的价值

数据集大小 特征尺寸		211 L6		043 20		52 26		00 83		330 L7		310 4 7		012 5 4
数据集	黑金		公元前		<i>t</i> :	在 应收则		收账款	ア 上海		画外音∗		明尼苏达州·	
原始特征(X) 突出 特点 (s)	✓	√	 	√	 	√	 	√	√	√	 	√	 	√
物流注册 随机森林 增强型 轻量级GBM CatBoost	0.907 0.891 0.929 0.939 0.925 0.915	0.918 0.902 0.938 0.942 0.937 0.923	0.892 0.879 0.906 0.910 0.912 0.892	0.902 0.899 0.912 0.915 0.919 0.902	0.862 0.850 0.870 0.887 0.879 0.902	0.869 0.885 0.904 0.903 0.899 0.912	0.916 0.809 0.824 0.821 0.825 0.903	0.922 0.846 0.843 0.831 0.877 0.904	0.870 0.929 0.925 0.932 0.931 0.887	0.882 0.933 0.931 0.944 0.942 0.910	0.539 0.663 0.690 0.679 0.664 0.631	0.551 0.672 0.693 0.686 0.682 0.642	0.899 0.938 0.958 0.952 0.956 0.939	0.921 0.945 0.964 0.963 0.968 0.948
维梅 塔格网 标签转换器 圣 雷康塔	0.766 0.918 0.913 0.933 0.929	- - - -	0.510 0.796 0.817 0.847 0.913	- - - -	0.653 0.521 0.700 0.941 0.907	- - - -	0.610 0.541 0.868 0.910 0.918	- - - -	0.744 0.914 0.927 0.931 0.931	- - - -	0.623 0.568 0.580 0.701 0.680	- - - -	0.958 0.968 0.887 0.977 0.968	- - - -
SwitchTab(自支持) 切换标签	0.917 0.942	-	0.903 0.923	-	0.900 0.928	-	0.904 0.922	-	0.931 0.958	-	0.629 0.708	-	0.969 0.982	-

表 2:不同方法在分类任务上的比较。对于每种方法,我们报告三个类别:1)仅原始特征,2)仅显著特征,3)使用显著特征的即插即用。最佳结果以粗体显示。添加的列·是多分类任务,报告准确率。二分类任务的其他结果用AUC来评估。

数据集	黑金	公元ī	前在	应收见	账 款 海	画外音·明尼苏达州·		
切换标签 (无切换)	0.918	0.909	0.902	0.896	0.912	0.689	0.968	
切换标签	0.942	0.923	0.928	0.922	0.958	0.708	0.982	

表 3:关于切换过程的模型性能消融。添加的列·是多分类任务,报告其准确率。其他二分类任务的结果用AUC来评估。

以原始数据作为附加特征。值得注意的是,所有传统方法的性能都可以得到提升,从而改善评估指标(例如 AUC)0。5%到3。5%(绝对差异)在各种数据集上,如表 2深灰色列所示。同时,我们还报告了在补充中仅使用显着特征作为输入时的结果。虽然改进相对较小,但符合我们的预期。在这种情况下缺乏相互信息会导致性能提升不那么显着。

可视化与讨论

在本节中,我们使用专为二元分类任务设计的 BK 数据集可视化 SwitchTab 学习到的特征。在预训练之后,我们将一个类的数据输入到第一个批次,将另一个类的数据输入到第二个批次,然后可视化相应的特征向量。如图 4 所示,嵌入 光和光2SwitchTab 中的数据虽然是从两个不同的类中提取出来的,但彼此之间有很大的重叠。这证实了相互信息是可以切换的。然而,显著的特征sī以及突出的特点sī截然不同,在捕捉每个类别的独特属性方面起着主导作用,并决定性地贡献

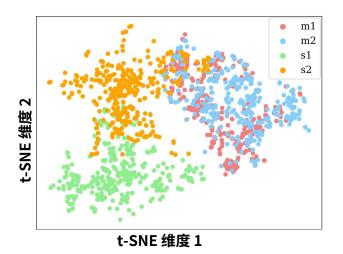


图 4: 二维空间中相互特征和显著特征的 t-SNE 可视化。

到分类边界。

消融研究

在本节中,我们将研究 SwitchTab 的关键模块,包括 切换过程的重要性和特征损坏率。我们使用表 2 中的所有 数据集,并采用相同的数据预处理和优化策略。

切换过程的贡献。为了证明所提模型的卓越性能直接源于关键的切换过程,我们报告了重建和未重建切换对的连接特征的结果,即(s_1 , H_2)和(s_2 , H_3),保持-

比率	0.0	0.1	0.2	0.3	0.4	0.5	0.6
黑金	0.927	0.938 0.9	400.942	0.932 0.	903 0.89	8	
公元前	0.911 (.9200.9	23 0.923	0.917 0.	910 0.90	2	
在	0.916 (0.922 0.9	250.928	0.927 0.	920 0.91	3	
应收账	約0.913 (0.915 0.9	18 0.922	20.9250.	920 0.91	4	
	0.683	0.694 0.6	99 0.708	0.947 0.9 30.7090. 9820.978	700 0.69	2	

表 4:特征损坏率的消融。多类分类任务·报告准确率。其他二分类任务用 AUC 来评估。

在所有实验中将特征损坏率设为 0.3。值得注意的是,如果没有切换机制,该框架会退化为更简单的自动编码器结构,并导致表 3 中的评估指标(例如 AUC)明显下降。

特征腐败率。我们还在表 4 中探讨了最佳特征损坏率。通过广泛的分析,我们发现最佳损坏率约为 0.3。因此,我们采用此值作为所有先前报告的实验的默认值。但是,必须强调的是,所选的比率可能并非每个数据集都始终最佳。我们还观察到,具有更高特征维度的数据集(例如 AR 或 VO)往往会受益于更大的损坏率,因为它们更有可能具有冗余特征。这一观察结果与先前关于表格数据的结论一致(Grinsztajn、Oyallon 和 Varoquaux 2022)。相反,对于具有低维特征的数据集(例如 BC),较小的损坏率也可以在我们的实验中产生更好的结果。

结论

受表征学习在计算视觉和自然语言处理领域取得的巨大成功的启发,我们希望将这一成功扩展到表格数据领域。与其他从对比学习角度解决这一问题的相关研究不同,我们引入了 SwitchTab,这是一个从生成模型的角度进行表征学习的新型预训练框架。从 SwitchTab 中学习到的嵌入不仅可以在下游任务上取得优异的表现,而且还代表了一个可区分的显着特征空间,可以作为即插即用的嵌入来增强更广泛的传统方法。我们坚信这项工作是朝着实现更具代表性、可解释性和结构化的表格数据表示迈出的关键一步。

参考

Abid, A.;和 Zou, JY 2019。对比变分自动编码器增强了显著特征。 $\hat{\nu}$ 文集,绝对/1902.04601。

Arik, S. Ö.; 和 Pfister, T. 2021。Tabnet: 专注的可解释表格学习。在*AAAI 人工智能会议论文集*,第 35 卷,6679-6687。

Asuncion, A.; 和 Newman, D. 2007。UCI 机器学习存储库。

Badirli, S.; Liu, X.; Xing, Z.; Bhowmik, A.; Doan, K.; 以及 Keerthi, SS 2020。梯度增强神经网络: Grownet。 *arXiv 预印本 arXiv:2002.07971*。

Bahri, D.; Jiang, H.; Tay, Y.; 和 Metzler, D. 2021. Scarf: 使用随机特征损坏的自监督对比学习。 *arXiv 预印本 arXiv:2106.15147*。

Baldi, P.; Sadowski, P.; 和 Whiteson, D. 2014. 利用深度学习寻找高能物理中的奇异粒子。*自然通讯*, 5(1): 4308。

Bengio, Y.; Courville, A.; 和 Vincent, P. 2013。表征学习:回顾与新视角。*IEEE 模式分析与机器智能交易*《自然》杂志 2007 年11 期,第 35 卷,第 8 期,1798-1828 页。

Bertin-Mahieux, T.; Ellis, DP; Whitman, B.; 以及 Lamere, P. 2011。百万首歌曲数据集。*学术共同体*。

Blackard, JA; 和 Dean, DJ 1999。人工神经网络和判别分析在根据制图变量预测森林覆盖类型方面的比较精度。农业中的计算机和电子产品《细胞与分子生物学杂志》, 24(3): 131-151。

Borisov, V.; Leemann, T.; Seßler, K.; Haug, J.; Pawelczyk, M.; 以及 Kasneci, G. 2022。深度神经网络和表格数据: 一项调查。 *IEEE 神经网络和学习系统学报*。

Bousmalis, K.; Trigeorgis, G.; Silberman, N.; Krishnan, D.; 以及 Erhan, D., 2016。域分离网络。*神经信息处理系统的进展*, 29。

Breiman, L. 2001.随机森林。*机器学习*, 45: 5–32。Breiman, L. 2017 年。*分类和回归树*. 劳特利奇。

Chapelle, O.; Chang, Y. 2011。Yahoo! 学习排名挑战概述。在 *学习排序挑战赛论文集*,1-24。PMLR。

Chen, S. 2020.贝叶斯二项式可靠性演示测试设计的一些最新进展。*USF 坦帕研究生论文和学位论文*。

Chen, S.; Kearns, WD; Fozard, JL; 和 Li, M. 2017。个性化跌倒风险评估,改善长期护理服务。在*2017 年度可靠性和可维护性研讨会 (RAMS)*,1-7。IEEE。

Chen, S.; Kong, N.; Sun, X.; Meng, H.; and Li, M. 2019. 具有潜在异质性的医院再入院时间风险索赔数据驱动建模。 *健康保健管理科学*《细胞与分子生物学杂志》, 22: 156–179。

Chen, S.; Lu, L.; 和 Li, M. 2017.多状态可靠性演示测试。*质量工程*《细胞与分子生物学杂志》, 29(3): 431-445。

Chen, S.; Lu, L.; Xiang, Y.; Lu, Q.; and Li, M. 2018. 一种用于老化基础设施中氯化物引起腐蚀的现场预评估的数据异质性建模和量化方法。 *可靠性工程与系统安全*,171:123–135。

Chen, S.; Lu, L.; Zhang, Q.; and Li, M. 2020a. 验收决策不确定性下的最佳二项式可靠性验证试验设计。质量工程《细胞与分子生物学》, 32(3): 492–508。

Chen, S.; Wu, J.; Hovakimyan, N.; 和 Yao, H. 2023. ReConTab: 针对表格数据的正则化对比表示学习。 *arXiv 预印本 arXiv:2310.18541*。

Chen, T.;以及 Guestrin, C. 2016。Xgboost:可扩展的树提升系统。在第 22 届 acm sigkdd 知识发现和数据挖掘国际会议论文集,785-794。

Chen, T.; Kornblith, S.; Norouzi, M.; 和 Hinton, G. 2020b。 视觉表征对比学习的简单框架。在 *国际机器学习会议*,1597– 1607 年。PMLR。

Chen, X.; Fan, H.; Girshick, R.; He, K. 2020c. 通过动量对比学习改进基线。 *arXiv 预印本 arXiv:2003.04297*。

- Chen, X.; He, K. 2021. 探索简单的暹罗表示学习。在/EEE/CVF 计算机视觉和模式识别会议论文集,15750-15758。
- Devlin, J.; Chang, M.-W.; Lee, K.; 以及 Toutanova, K. 2018。Bert: 用于语言理解的深度双向转换器的预训练。*arXiv* 预印本 arXiv:1810.04805。
- Ericsson, L.; Gouk, H.; Loy, CC; 以及 Hospedales, TM 2022。自监督表征学习:简介、进展和挑战。*IEEE 信号处理杂志*, 39(3): 42-62。
- Feichtenhofer, C.; Li, Y.; He, K.; 等人。2022 年。掩蔽自动编码器作为时空学习者。*神经信息处理系统的进展*《细胞与分子生物学杂志》, 35: 35946-35958。
- Geusebroek, J.-M.; Burghouts, GJ; 以及 Smeulders, AW 2005。阿姆斯特丹物体图像库。*国际计算机视觉杂志*,61: 103–112。
- Gorishniy, Y.; Rubachev, I.; Khrulkov, V.; 和 Babenko, A. 2021。重新审视表格数据的深度学习模型。*神经信息处理系统的进展*, 34: 18932–18943。
- Grill, J.-B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; 等人,2020年。引导你自己的潜能 一种自监督学习的新方法。神经信息处理系统的进展《细胞与分子生物学杂志》,33: 21271–21284。
- Grinsztajn, L.; Oyallon, E.; 和 Varoquaux, G. 2022. 为什么基于树的模型在典型的表格数据上仍然优于深度学习? 神经信息处理系统的进展《细胞与分子生物学杂志》, 35: 507–520。
- Guyon, I.; Sun-Hosoya, L.; Boullé, M.; Escalante, HJ; Escalera, S.; Liu, Z.; Jajetic, D.; Ray, B.; Saeed, M.; Sebag, M.; Statnikov, A.; Tu, W.; 以及 Viegas, E. 2019a。2015-2018 年 AutoML 挑战赛系列分析。*自动机器学习*,Springer 关于机器学习挑战的系列文章。
- Guyon, I.; Sun-Hosoya, L.; Boullé, M.; Escalante, HJ; Escalera, S.; Liu, Z.; Jajetic, D.; Ray, B.; Saeed, M.; Sebag, M.; 等人。2019b。AutoML 挑战系列分析。*自动机器学习*, 177。
- Hastie, TJ; 和 Pregibon, D. 2017。广义线性模型。 *S 中的统计模型*, 195-247。劳特利奇。
- He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; 以及 Girshick, R. 2022。蒙版自动编码器是可扩展的视觉学习器。在*IEEE/CVF 计算机视觉和模式识别会议论文集*,16000-16009年。
- He, K.; Fan, H.; Wu, Y.; Xie, S.; 以及 Girshick, R. 2020。无监督视觉表征学习的动量对比。在 *IEEE/CVF 计算机视觉和模式识别会议论文集*,9729–9738。
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. 深度残差学习用于图像识别。 *IEEE 计算机视觉和模式识别会议论文集*,770-778。
- Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.; Glorot, X.; Botvinick, M.; Mohamed, S.; 和 Lerchner, A. 2016. beta-vae: 使用受限变分框架学习基本视觉概念。在 *学习表征国际会议*。
- Hinton, G.; Srivastava, N.; 和 Swersky, K. 2012。机器学习神经 网络讲座 6a 小批量梯度下降概述。*引用于*,14(8):2。
- Huang, X.; Khetan, A.; Cvitkovic, M.; 和 Karnin, Z. 2020. Tabtransformer:使用上下文嵌入的表格数据建模。*arXiv 预印本 arXiv:2012.06678*。

- Jiang, J.; Lu, X.; Zhao, L.; Dazeley, R.; 和 Wang, M. 2022. 3D 点云表示学习中的掩蔽自动编码器。 *arXiv 预印本 arXiv:2207.01545*。
- Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; and Liu, T.-Y. 2017. Lightgbm: 一种高效的梯度提升决策树。 *神经信息处理系统的进展*, 30。
- Ke, G.; Xu, Z.; Zhang, J.; Bian, J.; 和 Liu, T.-Y。2019 年。DeepGBM: 一种由 GBDT 提炼的用于在线预测任务的深度学习框架。在第25届ACM SIGKDD 知识发现与数据挖掘国际会议论文集、384-394。
- Ke, G.; Zhang, J.; Xu, Z.; Bian, J.; 和 Liu, T.-Y。2018 年。TabNN: 一种用于表格数据的通用神经网络解决方案。
- Kim, H.; Mnih, A. 2018。通过分解来解开。在*国际机器学习会议*, 2649-2658。PMLR。
- Kingma, DP;以及 Welling, M. 2013。自动编码变分贝叶斯。 arXiv 预印本 arXiv:1312.6114。
- Klambauer, G.; Unterthiner, T.; Mayr, A.; 以及 Hochreiter, S. 2017。自规范化神经网络。*神经信息处理系统的进展*, 30。
- Kohavi, R.; 等人。1996年。扩大朴素贝叶斯分类器的准确性: 决策树混合。在*克德*, 第 96 卷, 202-207。
- Kolesnikov, A.; Zhai, X.; 和 Beyer, L. 2019。重新审视自监督视觉表征学习。在*IEEE/CVF 计算机视觉和模式识别会议论文集*,1920-1929年。
- Kossen, J.; Band, N.; Lyle, C.; Gomez, AN; Rainforth, T.; 以及 Gal, Y. 2021 年。数据点之间的自注意力:超越深度学习中的单个输入输出对。*神经信息处理系统的进展*《细胞与分子生物学杂志》, 34: 28742-28756。
- Li, X.; Guo, Y.; 以及 Schuurmans, D., 2015 年。带有标签表示学习的半监督零样本分类。在*IEEE 国际计算机视觉会议论文集*,4211-4219。
- Liakos, KG; Busato, P.; Moshou, D.; Pearson, S.; 和 Bochtis, D. 2018。农业中的机器学习:回顾。*传感器*, 18(8): 2674。
- Liu, FT; Ting, KM; 和 Zhou, Z.-H. 2008. 隔离森林。在*2008年第* 八届IEEE数据挖掘国际会议, 413-422。IEEE。
- Manas, O.; Lacoste, A.; Giró-i Nieto, X.; Vazquez, D.; 和 Rodriguez, P. 2021。季节性对比:使用未经整理的遥感数据进行无监督预训练。在*IEEE/CVF 国际计算机视觉会议论文集*,9414–9423。
- Meng, Q.; Catchpoole, D.; Skillicom, D.; 和 Kennedy, PJ 2017。用于特征提取的关系自动编码器。在*2017年国际神经网络联合会议(IJCNN)*, 364–371。IEEE。
- Moro, S.; Cortez, P.; 和 Rita, P. 2014. 一种预测银行电话营销成功率的数据驱动方法。*决策支持系统*,62: 22-31。
- Ouk, J.; Dada, D.; 以及 Kang, KT 2018。电信客户流失。 Pace, RK; Barry, R., 1997。稀疏空间自回归。 *统计与概率快报*《细胞与分子生物学》, 33(3): 291-297。
- Park, T.; Zhu, J.-Y.; Wang, O.; Lu, J.; Shechtman, E.; Efros, A.; 和 Zhang, R. 2020. 将自动编码器替换为深度图像处理。*神经信息处理系统的进展*《细胞与分子生物学杂志》, 33: 7198–7211。

Popov, S.; Morozov, S.; 和 Babenko, A. 2019。用于表格数据深度学习的神经无知决策集成。*arXiv 预印本 arXiv:1909.06312*

Potdar, K.; Pardawala, TS; 和 Pai, CD 2017。神经网络分类器的分类变量编码技术的比较研究。*国际计算机应用杂志*,175(4):7-9。

Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, AV; 以及 Gulin, A. 2018。CatBoost: 具有分类特征的无偏提升。神经信息处理系统的进展, 31。

Qayyum, A.; Qadir, J.; Bilal, M.; 和 Al-Fuqaha, A. 2020。安全且强大的医疗保健机器学习: 一项调查。*IEEE 生物医学工程评论*, 14: 156–180。

Qin, T.;和Liu, T.-Y. 2013.介绍LETOR 4.0数据集。 *arXiv 预印本 arXiv:1306.2597*。

Reed, CJ; Gupta, R.; Li, S.; Brockman, S.; Funk, C.; Clipp, B.; Candido, S.; Uyttendaele, M.; 和 Darrell, T. 2022. Scalemae: 一种用于多尺度地理空间表示学习的尺度感知掩蔽自动编码器。 arXiv 预印本 arXiv:2212.14532。

Sakar, CO; Polat, SO; Katircioglu, M.; 和 Kastro, Y. 2019. 使用多层感知器和 LSTM 循环神经网络实时预测在线购物者的购买意向。 神经计算与应用、31: 6893-6908。

Somepalli, G.; Goldblum, M.; Schwarzschild, A.; Bruss, CB; 以及 Goldstein, T. 2021. Saint: 通过行注意和对比预训练改进表格数据的神经网络。 *arXiv 预印本 arXiv:2106.01342*。

森德比,CK; 赖子,T.; Maaløe, L.; 森德比,SK; 和 Winther, O. 2016。梯形变分自动编码器。*神经信息处理系统的 进展*,29。

Song, W.; Shi, C.; Xiao, Z.; Duan, Z.; Xu, Y.; Zhang, M.; 以及 Tang, J. 2019. Autoint: 通过自注意力神经网络进行自动特征交互学习。 *第 28 届 ACM 信息与知识管理国际会议论文集*,1161–1170年。

Tao, R.; Zhao, P.; Wu, J.; Martin, NF; Harrison, MT; Ferreira, C.; Kalantari, Z.; 以及 Hovakimyan, N. 2022。通过强化学习和模仿学习优化作物管理。*arXiv 预印本arXiv:2209.09991*。

Tschannen, M.; Bachem, O.; 和 Lucic, M. 2018。基于自动编码器的表示学习的最新进展。

arXiv 预印本
arXiv:1812.05069。

Wang, B.; Lu, L.; Chen, S.; and Li, M. 2023a. 多阶段验收不确定性下的可靠性验证最优试验设计。质量工程, 0(0): 1–14。

Wang, R.; Shivanna, R.; Cheng, D.; Jain, S.; Lin, D.; Hong, L.; 和 Chi, E. 2021. Dcn v2: 改进的深度和交叉网络以及用于网络规模学习排名系统的实践课程。在*2021 年网络会议论文集*,1785–1797年。

Wang, Y.; Wu, J.; Hovakimyan, N.; 以及 Sun, R. 2023b。稀疏 GAN 的平衡训练。*第三十七届神经信息处理系统会议*。

Wright, RE 1995.逻辑回归。

Wu, J.; Hobbs, J.; 和 Hovakimyan, N. 2023。幻觉提高了无监督视觉表征学习的表现。在*IEEE/CVF 国际计算机视觉会议论文集*,16132–16143。

Wu, J.; Hovakimyan, N.; 和 Hobbs, J. 2023. Genco: 一种来自对比学习的辅助生成器,用于增强遥感中的小样本学习。 *arXiv* 预印本 arXiv:2307.14612。

Wu, J.; Pichler, D.; Marley, D.; Wilson, D.; Hovakimyan, N.; 和 Hobbs, J. 2023。扩展农业视觉:用于农业模式分析的大型航空图像数据集的扩展。 *arXiv 预印本 arXiv:2303.02460*。

Wu, J.; Tao, R.; Zhao, P.; Martin, NF; 和 Hovakimyan, N. 2022。通过深度强化学习和作物模拟优化氮管理。在*IEEE/CVF* 计算机视觉和模式识别会议论文集,1712-1720年。

Xiao, H.; Rasul, K.; 和 Vollgraf, R. 2017. Fashion-mnist: 用于对机器学习算法进行基准测试的新型图像数据集。 *arXiv 预印本 arXiv:1708.07747*。

Yin, P.; Neubig, G.; Yih, W.-t.; 以及 Riedel, S. 2020. TaBERT: 用于联合理解文本和表格数据的预训练。 *arXiv 预印本 arXiv:2005.08314*。

Yoon, J.; Zhang, Y.; Jordon, J.; 和 van der Schaar, M. 2020。Vime:将自我和半监督学习的成功扩展到表格领域。*神 经信息处理系统的进展*, 33: 11033–11043。

Yuan, G.-X.; Ho, C.-H.; 和 Lin, C.-J. 2011. 一种用于 l1 正则化逻辑回归的改进 glmnet。在*第 17 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集*,33–41。

Zhang, H.; Wang, M.; Liu, Y.; 和 Yuan, Y. 2020. FDN: 用于头部 姿势估计的特征解耦网络。在*AAAI 人工智能会议论文集*,第 34 卷,12789–12796。

Zhu, L.; Yu, FR; Wang, Y.; Ning, B.; 和 Tang, T. 2018. 智能交通系统中的大数据分析: 一项调查。 *IEEE 智能交通系统汇刊*, 20(1): 383-398。